wrangle_act

June 29, 2018

```
In [3]: import pandas as pd
        import numpy as np
        import requests
        import tweepy
        import json
        import re
        import matplotlib.pyplot as plt
        import warnings
```

1 Gather

```
In [4]: df_twitter = pd.read_csv('twitter-archive-enhanced.csv', encoding='utf-8')
        df_twitter
```

```
Out[4]:              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
        0    892420643555336193                    NaN                  NaN
        1    892177421306343426                    NaN                  NaN
        2    891815181378084864                    NaN                  NaN
        3    891689557279858688                    NaN                  NaN
        4    891327558926688256                    NaN                  NaN
        5    891087950875897856                    NaN                  NaN
        6    890971913173991426                    NaN                  NaN
        7    890729181411237888                    NaN                  NaN
        8    890609185150312448                    NaN                  NaN
        9    890240255349198849                    NaN                  NaN
        10   890006608113172480                    NaN                  NaN
        11   889880896479866881                    NaN                  NaN
        12   889665388333682689                    NaN                  NaN
        13   889638837579907072                    NaN                  NaN
        14   889531135344209921                    NaN                  NaN
        15   889278841981685760                    NaN                  NaN
        16   888917238123831296                    NaN                  NaN
        17   888804989199671297                    NaN                  NaN
        18   888554962724278272                    NaN                  NaN
        19   888202515573088257                    NaN                  NaN
        20   888078434458587136                    NaN                  NaN
        21   887705289381826560                    NaN                  NaN
```

```
22      887517139158093824                      NaN                    NaN
23      887473957103951883                      NaN                    NaN
24      887343217045368832                      NaN                    NaN
25      887101392804085760                      NaN                    NaN
26      886983233522544640                      NaN                    NaN
27      886736880519319552                      NaN                    NaN
28      886680336477933568                      NaN                    NaN
29      886366144734445568                      NaN                    NaN
...                    ...                      ...                    ...
2326    666411507551481857                      NaN                    NaN
2327    666407126856765440                      NaN                    NaN
2328    666396247373291520                      NaN                    NaN
2329    666373753744588802                      NaN                    NaN
2330    666362758909284353                      NaN                    NaN
2331    666353288456101888                      NaN                    NaN
2332    666345417576210432                      NaN                    NaN
2333    666337882303524864                      NaN                    NaN
2334    666293911632134144                      NaN                    NaN
2335    666287406224695296                      NaN                    NaN
2336    666273097616637952                      NaN                    NaN
2337    666268910803644416                      NaN                    NaN
2338    666104133288665088                      NaN                    NaN
2339    666102155909144576                      NaN                    NaN
2340    666099513787052032                      NaN                    NaN
2341    666094000022159362                      NaN                    NaN
2342    666082916733198337                      NaN                    NaN
2343    666073100786774016                      NaN                    NaN
2344    666071193221509120                      NaN                    NaN
2345    666063827256086533                      NaN                    NaN
2346    666058600524156928                      NaN                    NaN
2347    666057090499244032                      NaN                    NaN
2348    666055525042405380                      NaN                    NaN
2349    666051853826850816                      NaN                    NaN
2350    666050758794694657                      NaN                    NaN
2351    666049248165822465                      NaN                    NaN
2352    666044226329800704                      NaN                    NaN
2353    666033412701032449                      NaN                    NaN
2354    666029285002620928                      NaN                    NaN
2355    666020888022790149                      NaN                    NaN

                        timestamp  \
0       2017-08-01 16:23:56 +0000
1       2017-08-01 00:17:27 +0000
2       2017-07-31 00:18:03 +0000
3       2017-07-30 15:58:51 +0000
4       2017-07-29 16:00:24 +0000
5       2017-07-29 00:08:17 +0000
6       2017-07-28 16:27:12 +0000
```

```
7      2017-07-28 00:22:40 +0000
8      2017-07-27 16:25:51 +0000
9      2017-07-26 15:59:51 +0000
10     2017-07-26 00:31:25 +0000
11     2017-07-25 16:11:53 +0000
12     2017-07-25 01:55:32 +0000
13     2017-07-25 00:10:02 +0000
14     2017-07-24 17:02:04 +0000
15     2017-07-24 00:19:32 +0000
16     2017-07-23 00:22:39 +0000
17     2017-07-22 16:56:37 +0000
18     2017-07-22 00:23:06 +0000
19     2017-07-21 01:02:36 +0000
20     2017-07-20 16:49:33 +0000
21     2017-07-19 16:06:48 +0000
22     2017-07-19 03:39:09 +0000
23     2017-07-19 00:47:34 +0000
24     2017-07-18 16:08:03 +0000
25     2017-07-18 00:07:08 +0000
26     2017-07-17 16:17:36 +0000
27     2017-07-16 23:58:41 +0000
28     2017-07-16 20:14:00 +0000
29     2017-07-15 23:25:31 +0000
...                     ...
2326   2015-11-17 00:24:19 +0000
2327   2015-11-17 00:06:54 +0000
2328   2015-11-16 23:23:41 +0000
2329   2015-11-16 21:54:18 +0000
2330   2015-11-16 21:10:36 +0000
2331   2015-11-16 20:32:58 +0000
2332   2015-11-16 20:01:42 +0000
2333   2015-11-16 19:31:45 +0000
2334   2015-11-16 16:37:02 +0000
2335   2015-11-16 16:11:11 +0000
2336   2015-11-16 15:14:19 +0000
2337   2015-11-16 14:57:41 +0000
2338   2015-11-16 04:02:55 +0000
2339   2015-11-16 03:55:04 +0000
2340   2015-11-16 03:44:34 +0000
2341   2015-11-16 03:22:39 +0000
2342   2015-11-16 02:38:37 +0000
2343   2015-11-16 01:59:36 +0000
2344   2015-11-16 01:52:02 +0000
2345   2015-11-16 01:22:45 +0000
2346   2015-11-16 01:01:59 +0000
2347   2015-11-16 00:55:59 +0000
2348   2015-11-16 00:49:46 +0000
2349   2015-11-16 00:35:11 +0000
```

```
2350   2015-11-16 00:30:50 +0000
2351   2015-11-16 00:24:50 +0000
2352   2015-11-16 00:04:52 +0000
2353   2015-11-15 23:21:54 +0000
2354   2015-11-15 23:05:30 +0000
2355   2015-11-15 22:32:08 +0000

                                                        source  \
0      <a href="http://twitter.com/download/iphone" r...
1      <a href="http://twitter.com/download/iphone" r...
2      <a href="http://twitter.com/download/iphone" r...
3      <a href="http://twitter.com/download/iphone" r...
4      <a href="http://twitter.com/download/iphone" r...
5      <a href="http://twitter.com/download/iphone" r...
6      <a href="http://twitter.com/download/iphone" r...
7      <a href="http://twitter.com/download/iphone" r...
8      <a href="http://twitter.com/download/iphone" r...
9      <a href="http://twitter.com/download/iphone" r...
10     <a href="http://twitter.com/download/iphone" r...
11     <a href="http://twitter.com/download/iphone" r...
12     <a href="http://twitter.com/download/iphone" r...
13     <a href="http://twitter.com/download/iphone" r...
14     <a href="http://twitter.com/download/iphone" r...
15     <a href="http://twitter.com/download/iphone" r...
16     <a href="http://twitter.com/download/iphone" r...
17     <a href="http://twitter.com/download/iphone" r...
18     <a href="http://twitter.com/download/iphone" r...
19     <a href="http://twitter.com/download/iphone" r...
20     <a href="http://twitter.com/download/iphone" r...
21     <a href="http://twitter.com/download/iphone" r...
22     <a href="http://twitter.com/download/iphone" r...
23     <a href="http://twitter.com/download/iphone" r...
24     <a href="http://twitter.com/download/iphone" r...
25     <a href="http://twitter.com/download/iphone" r...
26     <a href="http://twitter.com/download/iphone" r...
27     <a href="http://twitter.com/download/iphone" r...
28     <a href="http://twitter.com/download/iphone" r...
29     <a href="http://twitter.com/download/iphone" r...
...                                                  ...
2326   <a href="http://twitter.com/download/iphone" r...
2327   <a href="http://twitter.com/download/iphone" r...
2328   <a href="http://twitter.com/download/iphone" r...
2329   <a href="http://twitter.com/download/iphone" r...
2330   <a href="http://twitter.com/download/iphone" r...
2331   <a href="http://twitter.com/download/iphone" r...
2332   <a href="http://twitter.com/download/iphone" r...
2333   <a href="http://twitter.com/download/iphone" r...
2334   <a href="http://twitter.com/download/iphone" r...
```

```
2335  <a href="http://twitter.com/download/iphone" r...
2336  <a href="http://twitter.com/download/iphone" r...
2337  <a href="http://twitter.com/download/iphone" r...
2338  <a href="http://twitter.com/download/iphone" r...
2339  <a href="http://twitter.com/download/iphone" r...
2340  <a href="http://twitter.com/download/iphone" r...
2341  <a href="http://twitter.com/download/iphone" r...
2342  <a href="http://twitter.com/download/iphone" r...
2343  <a href="http://twitter.com/download/iphone" r...
2344  <a href="http://twitter.com/download/iphone" r...
2345  <a href="http://twitter.com/download/iphone" r...
2346  <a href="http://twitter.com/download/iphone" r...
2347  <a href="http://twitter.com/download/iphone" r...
2348  <a href="http://twitter.com/download/iphone" r...
2349  <a href="http://twitter.com/download/iphone" r...
2350  <a href="http://twitter.com/download/iphone" r...
2351  <a href="http://twitter.com/download/iphone" r...
2352  <a href="http://twitter.com/download/iphone" r...
2353  <a href="http://twitter.com/download/iphone" r...
2354  <a href="http://twitter.com/download/iphone" r...
2355  <a href="http://twitter.com/download/iphone" r...


                                                   text  retweeted_status_id  \
0      This is Phineas. He's a mystical boy. Only eve...                  NaN
1      This is Tilly. She's just checking pup on you...                  NaN
2      This is Archie. He is a rare Norwegian Pouncin...                  NaN
3      This is Darla. She commenced a snooze mid meal...                  NaN
4      This is Franklin. He would like you to stop ca...                  NaN
5      Here we have a majestic great white breaching ...                  NaN
6      Meet Jax. He enjoys ice cream so much he gets ...                  NaN
7      When you watch your owner call another dog a g...                  NaN
8      This is Zoey. She doesn't want to be one of th...                  NaN
9      This is Cassie. She is a college pup. Studying...                  NaN
10     This is Koda. He is a South Australian decksha...                  NaN
11     This is Bruno. He is a service shark. Only get...                  NaN
12     Here's a puppo that seems to be on the fence a...                  NaN
13     This is Ted. He does his best. Sometimes that'...                  NaN
14     This is Stuart. He's sporting his favorite fan...                  NaN
15     This is Oliver. You're witnessing one of his m...                  NaN
16     This is Jim. He found a fren. Taught him how t...                  NaN
17     This is Zeke. He has a new stick. Very proud o...                  NaN
18     This is Ralphus. He's powering up. Attempting ...                  NaN
19     RT @dog_rates: This is Canela. She attempted s...         8.874740e+17
20     This is Gerald. He was just told he didn't get...                  NaN
21     This is Jeffrey. He has a monopoly on the pool...                  NaN
22     I've yet to rate a Venezuelan Hover Wiener. Th...                  NaN
23     This is Canela. She attempted some fancy porch...                  NaN
24     You may not have known you needed to see this ...                  NaN
```

5

```
25     This... is a Jubilant Antarctic House Bear. We...              NaN
26     This is Maya. She's very shy. Rarely leaves he...             NaN
27     This is Mingus. He's a wonderful father to his...            NaN
28     This is Derek. He's late for a dog meeting. 13...            NaN
29     This is Roscoe. Another pupper fallen victim t...            NaN
...                                                 ...              ...
2326   This is quite the dog. Gets really excited whe...            NaN
2327   This is a southern Vesuvius bumblegruff. Can d...            NaN
2328   Oh goodness. A super rare northeast Qdoba kang...            NaN
2329   Those are sunglasses and a jean jacket. 11/10 ...            NaN
2330   Unique dog here. Very small. Lives in containe...            NaN
2331   Here we have a mixed Asiago from the Galápagos...            NaN
2332   Look at this jokester thinking seat belt laws ...            NaN
2333   This is an extremely rare horned Parthenon. No...            NaN
2334   This is a funny dog. Weird toes. Won't come do...            NaN
2335   This is an Albanian 3 1/2 legged  Episcopalian...            NaN
2336       Can take selfies 11/10 https://t.co/ws2AMaNwPW           NaN
2337   Very concerned about fellow dog trapped in com...            NaN
2338   Not familiar with this breed. No tail (weird)...             NaN
2339   Oh my. Here you are seeing an Adobe Setter giv...            NaN
2340   Can stand on stump for what seems like a while...            NaN
2341   This appears to be a Mongolian Presbyterian mi...            NaN
2342   Here we have a well-established sunblockerspan...            NaN
2343   Let's hope this flight isn't Malaysian (lol). ...            NaN
2344   Here we have a northern speckled Rhododendron...             NaN
2345   This is the happiest dog you will ever see. Ve...            NaN
2346   Here is the Rand Paul of retrievers folks! He'...            NaN
2347   My oh my. This is a rare blond Canadian terrie...            NaN
2348   Here is a Siberian heavily armored polar bear ...            NaN
2349   This is an odd dog. Hard on the outside but lo...            NaN
2350   This is a truly beautiful English Wilson Staff...            NaN
2351   Here we have a 1949 1st generation vulpix. Enj...            NaN
2352   This is a purebred Piers Morgan. Loves to Netf...            NaN
2353   Here is a very happy pup. Big fan of well-main...            NaN
2354   This is a western brown Mitsubishi terrier. Up...            NaN
2355   Here we have a Japanese Irish Setter. Lost eye...            NaN

       retweeted_status_user_id retweeted_status_timestamp  \
0                           NaN                         NaN
1                           NaN                         NaN
2                           NaN                         NaN
3                           NaN                         NaN
4                           NaN                         NaN
5                           NaN                         NaN
6                           NaN                         NaN
7                           NaN                         NaN
8                           NaN                         NaN
9                           NaN                         NaN
```

10	NaN	NaN
11	NaN	NaN
12	NaN	NaN
13	NaN	NaN
14	NaN	NaN
15	NaN	NaN
16	NaN	NaN
17	NaN	NaN
18	NaN	NaN
19	4.196984e+09	2017-07-19 00:47:34 +0000
20	NaN	NaN
21	NaN	NaN
22	NaN	NaN
23	NaN	NaN
24	NaN	NaN
25	NaN	NaN
26	NaN	NaN
27	NaN	NaN
28	NaN	NaN
29	NaN	NaN
...	...	...
2326	NaN	NaN
2327	NaN	NaN
2328	NaN	NaN
2329	NaN	NaN
2330	NaN	NaN
2331	NaN	NaN
2332	NaN	NaN
2333	NaN	NaN
2334	NaN	NaN
2335	NaN	NaN
2336	NaN	NaN
2337	NaN	NaN
2338	NaN	NaN
2339	NaN	NaN
2340	NaN	NaN
2341	NaN	NaN
2342	NaN	NaN
2343	NaN	NaN
2344	NaN	NaN
2345	NaN	NaN
2346	NaN	NaN
2347	NaN	NaN
2348	NaN	NaN
2349	NaN	NaN
2350	NaN	NaN
2351	NaN	NaN
2352	NaN	NaN

```
2353                        NaN                    NaN
2354                        NaN                    NaN
2355                        NaN                    NaN


                                    expanded_urls   rating_numerator  \
0      https://twitter.com/dog_rates/status/892420643...               13
1      https://twitter.com/dog_rates/status/892177421...               13
2      https://twitter.com/dog_rates/status/891815181...               12
3      https://twitter.com/dog_rates/status/891689557...               13
4      https://twitter.com/dog_rates/status/891327558...               12
5      https://twitter.com/dog_rates/status/891087950...               13
6      https://gofundme.com/ydvmve-surgery-for-jax,ht...               13
7      https://twitter.com/dog_rates/status/890729181...               13
8      https://twitter.com/dog_rates/status/890609185...               13
9      https://twitter.com/dog_rates/status/890240255...               14
10     https://twitter.com/dog_rates/status/890006608...               13
11     https://twitter.com/dog_rates/status/889880896...               13
12     https://twitter.com/dog_rates/status/889665388...               13
13     https://twitter.com/dog_rates/status/889638837...               12
14     https://twitter.com/dog_rates/status/889531135...               13
15     https://twitter.com/dog_rates/status/889278841...               13
16     https://twitter.com/dog_rates/status/888917238...               12
17     https://twitter.com/dog_rates/status/888804989...               13
18     https://twitter.com/dog_rates/status/888554962...               13
19     https://twitter.com/dog_rates/status/887473957...               13
20     https://twitter.com/dog_rates/status/888078434...               12
21     https://twitter.com/dog_rates/status/887705289...               13
22     https://twitter.com/dog_rates/status/887517139...               14
23     https://twitter.com/dog_rates/status/887473957...               13
24     https://twitter.com/dog_rates/status/887343217...               13
25     https://twitter.com/dog_rates/status/887101392...               12
26     https://twitter.com/dog_rates/status/886983233...               13
27     https://www.gofundme.com/mingusneedsus,https:/...               13
28     https://twitter.com/dog_rates/status/886680336...               13
29     https://twitter.com/dog_rates/status/886366144...               12
...                                             ...              ...
2326   https://twitter.com/dog_rates/status/666411507...                2
2327   https://twitter.com/dog_rates/status/666407126...                7
2328   https://twitter.com/dog_rates/status/666396247...                9
2329   https://twitter.com/dog_rates/status/666373753...               11
2330   https://twitter.com/dog_rates/status/666362758...                6
2331   https://twitter.com/dog_rates/status/666353288...                8
2332   https://twitter.com/dog_rates/status/666345417...               10
2333   https://twitter.com/dog_rates/status/666337882...                9
2334   https://twitter.com/dog_rates/status/666293911...                3
2335   https://twitter.com/dog_rates/status/666287406...                1
2336   https://twitter.com/dog_rates/status/666273097...               11
2337   https://twitter.com/dog_rates/status/666268910...               10
```

```
2338  https://twitter.com/dog_rates/status/666104133...                1
2339  https://twitter.com/dog_rates/status/666102155...               11
2340  https://twitter.com/dog_rates/status/666099513...                8
2341  https://twitter.com/dog_rates/status/666094000...                9
2342  https://twitter.com/dog_rates/status/666082916...                6
2343  https://twitter.com/dog_rates/status/666073100...               10
2344  https://twitter.com/dog_rates/status/666071193...                9
2345  https://twitter.com/dog_rates/status/666063827...               10
2346  https://twitter.com/dog_rates/status/666058600...                8
2347  https://twitter.com/dog_rates/status/666057090...                9
2348  https://twitter.com/dog_rates/status/666055525...               10
2349  https://twitter.com/dog_rates/status/666051853...                2
2350  https://twitter.com/dog_rates/status/666050758...               10
2351  https://twitter.com/dog_rates/status/666049248...                5
2352  https://twitter.com/dog_rates/status/666044226...                6
2353  https://twitter.com/dog_rates/status/666033412...                9
2354  https://twitter.com/dog_rates/status/666029285...                7
2355  https://twitter.com/dog_rates/status/666020888...                8
```

	rating_denominator	name	doggo	floofer	pupper	puppo
0	10	Phineas	None	None	None	None
1	10	Tilly	None	None	None	None
2	10	Archie	None	None	None	None
3	10	Darla	None	None	None	None
4	10	Franklin	None	None	None	None
5	10	None	None	None	None	None
6	10	Jax	None	None	None	None
7	10	None	None	None	None	None
8	10	Zoey	None	None	None	None
9	10	Cassie	doggo	None	None	None
10	10	Koda	None	None	None	None
11	10	Bruno	None	None	None	None
12	10	None	None	None	None	puppo
13	10	Ted	None	None	None	None
14	10	Stuart	None	None	None	puppo
15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
19	10	Canela	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None

```
28                    10    Derek   None    None    None    None
29                    10   Roscoe   None    None  pupper    None
...                  ...      ...    ...     ...     ...     ...
2326                  10    quite   None    None    None    None
2327                  10        a   None    None    None    None
2328                  10     None   None    None    None    None
2329                  10     None   None    None    None    None
2330                  10     None   None    None    None    None
2331                  10     None   None    None    None    None
2332                  10     None   None    None    None    None
2333                  10       an   None    None    None    None
2334                  10        a   None    None    None    None
2335                   2       an   None    None    None    None
2336                  10     None   None    None    None    None
2337                  10     None   None    None    None    None
2338                  10     None   None    None    None    None
2339                  10     None   None    None    None    None
2340                  10     None   None    None    None    None
2341                  10     None   None    None    None    None
2342                  10     None   None    None    None    None
2343                  10     None   None    None    None    None
2344                  10     None   None    None    None    None
2345                  10      the   None    None    None    None
2346                  10      the   None    None    None    None
2347                  10        a   None    None    None    None
2348                  10        a   None    None    None    None
2349                  10       an   None    None    None    None
2350                  10        a   None    None    None    None
2351                  10     None   None    None    None    None
2352                  10        a   None    None    None    None
2353                  10        a   None    None    None    None
2354                  10        a   None    None    None    None
2355                  10     None   None    None    None    None

[2356 rows x 17 columns]

In [3]: url = "https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictic
        response = requests.get(url)

In [4]: with open('image_predictions.tsv', 'wb') as file:
            file.write(response.content)

In [5]: df_image = pd.read_csv('image_predictions.tsv', sep='\t')
        df_image

Out[5]:                 tweet_id                                            jpg_url  \
        0    666020888022790149    https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg
        1    666029285002620928    https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
        2    666033412701032449    https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
```

```
3     666044226329800704    https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
4     666049248165822465    https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg
5     666050758794694657    https://pbs.twimg.com/media/CT5Jof1WUAEuVxN.jpg
6     666051853826850816    https://pbs.twimg.com/media/CT5KoJ1WoAAJash.jpg
7     666055525042405380    https://pbs.twimg.com/media/CT5N9tpXIAAifs1.jpg
8     666057090499244032    https://pbs.twimg.com/media/CT5PY90WoAAQGLo.jpg
9     666058600524156928    https://pbs.twimg.com/media/CT5Qw94XAAA_2dP.jpg
10    666063827256086533    https://pbs.twimg.com/media/CT5Vg_wXIAAXfnj.jpg
11    666071193221509120    https://pbs.twimg.com/media/CT5cN_3WEAAlOoZ.jpg
12    666073100786774016    https://pbs.twimg.com/media/CT5d9DZXAAALcwe.jpg
13    666082916733198337    https://pbs.twimg.com/media/CT5m4VGWEAAtKc8.jpg
14    666094000022159362    https://pbs.twimg.com/media/CT5w9gUW4AAsBNN.jpg
15    666099513787052032    https://pbs.twimg.com/media/CT51-JJUEAA6hV8.jpg
16    666102155909144576    https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg
17    666104133288665088    https://pbs.twimg.com/media/CT56LSZWoAAlJj2.jpg
18    666268910803644416    https://pbs.twimg.com/media/CT8QCd1WEAADXws.jpg
19    666273097616637952    https://pbs.twimg.com/media/CT8T1mtUwAA3aqm.jpg
20    666287406224695296    https://pbs.twimg.com/media/CT8g3BpUEAAuFjg.jpg
21    666293911632134144    https://pbs.twimg.com/media/CT8mx7KW4AEQu8N.jpg
22    666337882303524864    https://pbs.twimg.com/media/CT9OwFIWEAMuRje.jpg
23    666345417576210432    https://pbs.twimg.com/media/CT9Vn7PWoAA_ZCM.jpg
24    666353288456101888    https://pbs.twimg.com/media/CT9cxOtUEAAhNN_.jpg
25    666362758909284353    https://pbs.twimg.com/media/CT9lXGsUcAAyUFt.jpg
26    666373753744588802    https://pbs.twimg.com/media/CT9vZEYWUAAlZO5.jpg
27    666396247373291520    https://pbs.twimg.com/media/CT-D2ZHWIAA3gK1.jpg
28    666407126856765440    https://pbs.twimg.com/media/CT-NvwmW4AAugGZ.jpg
29    666411507551481857    https://pbs.twimg.com/media/CT-RugiWIAELEaq.jpg
...            ...                                          ...
2045  886366144734445568    https://pbs.twimg.com/media/DE0BTnQUwAApKEH.jpg
2046  886680336477933568    https://pbs.twimg.com/media/DE4fEDzWAAAyHMM.jpg
2047  886736880519319552    https://pbs.twimg.com/media/DE5Se8FXcAAJFx4.jpg
2048  886983233522544640    https://pbs.twimg.com/media/DE8yicJWOAAAvBJ.jpg
2049  887101392804085760    https://pbs.twimg.com/media/DE-eAq6UwAA-jaE.jpg
2050  887343217045368832    https://pbs.twimg.com/ext_tw_video_thumb/88734...
2051  887473957103951883    https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2052  887517139158093824    https://pbs.twimg.com/ext_tw_video_thumb/88751...
2053  887705289381826560    https://pbs.twimg.com/media/DFHDQBbXgAEqY7t.jpg
2054  888078434458587136    https://pbs.twimg.com/media/DFMWn56WsAAkA7B.jpg
2055  888202515573088257    https://pbs.twimg.com/media/DFDw2tyUQAAAFke.jpg
2056  888554962724278272    https://pbs.twimg.com/media/DFTH_O-UQAACu20.jpg
2057  888804989199671297    https://pbs.twimg.com/media/DFWra-3VYAA2piG.jpg
2058  888917238123831296    https://pbs.twimg.com/media/DFYRgsOUQAARGhO.jpg
2059  889278841981685760    https://pbs.twimg.com/ext_tw_video_thumb/88927...
2060  889531135344209921    https://pbs.twimg.com/media/DFg_2PVWOAEHN3p.jpg
2061  889638837579907072    https://pbs.twimg.com/media/DFihzFfXsAYGDPR.jpg
2062  889665388333682689    https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg
2063  889880896479866881    https://pbs.twimg.com/media/DFl99B1WsAITKsg.jpg
2064  890006608113172480    https://pbs.twimg.com/media/DFnwSY4WAAAMliS.jpg
```

```
2065   890240255349198849    https://pbs.twimg.com/media/DFrEyVuW0AAO3t9.jpg
2066   890609185150312448    https://pbs.twimg.com/media/DFwUU__XcAEpyXI.jpg
2067   890729181411237888    https://pbs.twimg.com/media/DFyBahAVwAAhUTd.jpg
2068   890971913173991426    https://pbs.twimg.com/media/DF1eOmZXUAALUcq.jpg
2069   891087950875897856    https://pbs.twimg.com/media/DF3HwyEWsAABqE6.jpg
2070   891327558926688256    https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
2071   891689557279858688    https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
2072   891815181378084864    https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
2073   892177421306343426    https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
2074   892420643555336193    https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg

      img_num                         p1    p1_conf   p1_dog  \
0           1      Welsh_springer_spaniel   0.465074     True
1           1                     redbone   0.506826     True
2           1             German_shepherd   0.596461     True
3           1          Rhodesian_ridgeback   0.408143     True
4           1           miniature_pinscher   0.560311     True
5           1         Bernese_mountain_dog   0.651137     True
6           1                   box_turtle   0.933012    False
7           1                         chow   0.692517     True
8           1                shopping_cart   0.962465    False
9           1             miniature_poodle   0.201493     True
10          1             golden_retriever   0.775930     True
11          1                 Gordon_setter   0.503672     True
12          1                 Walker_hound   0.260857     True
13          1                          pug   0.489814     True
14          1                    bloodhound   0.195217     True
15          1                        Lhasa   0.582330     True
16          1                English_setter   0.298617     True
17          1                          hen   0.965932    False
18          1              desktop_computer   0.086502    False
19          1             Italian_greyhound   0.176053     True
20          1                  Maltese_dog   0.857531     True
21          1               three-toed_sloth   0.914671    False
22          1                           ox   0.416669    False
23          1             golden_retriever   0.858744     True
24          1                      malamute   0.336874     True
25          1                    guinea_pig   0.996496    False
26          1     soft-coated_wheaten_terrier   0.326467     True
27          1                     Chihuahua   0.978108     True
28          1         black-and-tan_coonhound   0.529139     True
29          1                         coho   0.404640    False
...        ...                          ...        ...      ...
2045        1                French_bulldog   0.999201     True
2046        1                  convertible   0.738995    False
2047        1                        kuvasz   0.309706     True
2048        2                     Chihuahua   0.793469     True
2049        1                       Samoyed   0.733942     True
```

2050	1	Mexican_hairless	0.330741	True
2051	2	Pembroke	0.809197	True
2052	1	limousine	0.130432	False
2053	1	basset	0.821664	True
2054	1	French_bulldog	0.995026	True
2055	2	Pembroke	0.809197	True
2056	3	Siberian_husky	0.700377	True
2057	1	golden_retriever	0.469760	True
2058	1	golden_retriever	0.714719	True
2059	1	whippet	0.626152	True
2060	1	golden_retriever	0.953442	True
2061	1	French_bulldog	0.991650	True
2062	1	Pembroke	0.966327	True
2063	1	French_bulldog	0.377417	True
2064	1	Samoyed	0.957979	True
2065	1	Pembroke	0.511319	True
2066	1	Irish_terrier	0.487574	True
2067	2	Pomeranian	0.566142	True
2068	1	Appenzeller	0.341703	True
2069	1	Chesapeake_Bay_retriever	0.425595	True
2070	2	basset	0.555712	True
2071	1	paper_towel	0.170278	False
2072	1	Chihuahua	0.716012	True
2073	1	Chihuahua	0.323581	True
2074	1	orange	0.097049	False

	p2	p2_conf	p2_dog	p3 \
0	collie	0.156665	True	Shetland_sheepdog
1	miniature_pinscher	0.074192	True	Rhodesian_ridgeback
2	malinois	0.138584	True	bloodhound
3	redbone	0.360687	True	miniature_pinscher
4	Rottweiler	0.243682	True	Doberman
5	English_springer	0.263788	True	Greater_Swiss_Mountain_dog
6	mud_turtle	0.045885	False	terrapin
7	Tibetan_mastiff	0.058279	True	fur_coat
8	shopping_basket	0.014594	False	golden_retriever
9	komondor	0.192305	True	soft-coated_wheaten_terrier
10	Tibetan_mastiff	0.093718	True	Labrador_retriever
11	Yorkshire_terrier	0.174201	True	Pekinese
12	English_foxhound	0.175382	True	Ibizan_hound
13	bull_mastiff	0.404722	True	French_bulldog
14	German_shepherd	0.078260	True	malinois
15	Shih-Tzu	0.166192	True	Dandie_Dinmont
16	Newfoundland	0.149842	True	borzoi
17	cock	0.033919	False	partridge
18	desk	0.085547	False	bookcase
19	toy_terrier	0.111884	True	basenji
20	toy_poodle	0.063064	True	miniature_poodle

21	otter	0.015250	False	great_grey_owl
22	Newfoundland	0.278407	True	groenendael
23	Chesapeake_Bay_retriever	0.054787	True	Labrador_retriever
24	Siberian_husky	0.147655	True	Eskimo_dog
25	skunk	0.002402	False	hamster
26	Afghan_hound	0.259551	True	briard
27	toy_terrier	0.009397	True	papillon
28	bloodhound	0.244220	True	flat-coated_retriever
29	barracouta	0.271485	False	gar
...	...	...	...	...
2045	Chihuahua	0.000361	True	Boston_bull
2046	sports_car	0.139952	False	car_wheel
2047	Great_Pyrenees	0.186136	True	Dandie_Dinmont
2048	toy_terrier	0.143528	True	can_opener
2049	Eskimo_dog	0.035029	True	Staffordshire_bullterrier
2050	sea_lion	0.275645	False	Weimaraner
2051	Rhodesian_ridgeback	0.054950	True	beagle
2052	tow_truck	0.029175	False	shopping_cart
2053	redbone	0.087582	True	Weimaraner
2054	pug	0.000932	True	bull_mastiff
2055	Rhodesian_ridgeback	0.054950	True	beagle
2056	Eskimo_dog	0.166511	True	malamute
2057	Labrador_retriever	0.184172	True	English_setter
2058	Tibetan_mastiff	0.120184	True	Labrador_retriever
2059	borzoi	0.194742	True	Saluki
2060	Labrador_retriever	0.013834	True	redbone
2061	boxer	0.002129	True	Staffordshire_bullterrier
2062	Cardigan	0.027356	True	basenji
2063	Labrador_retriever	0.151317	True	muzzle
2064	Pomeranian	0.013884	True	chow
2065	Cardigan	0.451038	True	Chihuahua
2066	Irish_setter	0.193054	True	Chesapeake_Bay_retriever
2067	Eskimo_dog	0.178406	True	Pembroke
2068	Border_collie	0.199287	True	ice_lolly
2069	Irish_terrier	0.116317	True	Indian_elephant
2070	English_springer	0.225770	True	German_short-haired_pointer
2071	Labrador_retriever	0.168086	True	spatula
2072	malamute	0.078253	True	kelpie
2073	Pekinese	0.090647	True	papillon
2074	bagel	0.085851	False	banana

	p3_conf	p3_dog
0	0.061428	True
1	0.072010	True
2	0.116197	True
3	0.222752	True
4	0.154629	True
5	0.016199	True

```
6      0.017885    False
7      0.054449    False
8      0.007959     True
9      0.082086     True
10     0.072427     True
11     0.109454     True
12     0.097471     True
13     0.048960     True
14     0.075628     True
15     0.089688     True
16     0.133649     True
17     0.000052    False
18     0.079480    False
19     0.111152     True
20     0.025581     True
21     0.013207    False
22     0.102643     True
23     0.014241     True
24     0.093412     True
25     0.000461    False
26     0.206803     True
27     0.004577     True
28     0.173810     True
29     0.189945    False
...      ...        ...
2045   0.000076     True
2046   0.044173    False
2047   0.086346     True
2048   0.032253    False
2049   0.029705     True
2050   0.134203     True
2051   0.038915     True
2052   0.026321    False
2053   0.026236     True
2054   0.000903     True
2055   0.038915     True
2056   0.111411     True
2057   0.073482     True
2058   0.105506     True
2059   0.027351     True
2060   0.007958     True
2061   0.001498     True
2062   0.004633     True
2063   0.082981    False
2064   0.008167     True
2065   0.029248     True
2066   0.118184     True
2067   0.076507     True
```

```
     2068   0.193548    False
     2069   0.076902    False
     2070   0.175219     True
     2071   0.040836    False
     2072   0.031379     True
     2073   0.068957     True
     2074   0.076110    False

     [2075 rows x 12 columns]

In [4]: consumer_key = 'Consumer Key'
        consumer_secret = 'Consumer Secret'
        access_token = 'Access Token'
        access_secret = 'Access Secret'

In [5]: auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
        auth.set_access_token(access_token, access_secret)
        api = tweepy.API(auth,parser = tweepy.parsers.JSONParser(), wait_on_rate_limit = True, w

In [6]: # Liste where we will store the dictionaries of our result
        df_list = []
        # Liste frame where we will store the tweet_id of the errors
        error_list = []


        # Get the tweet object for all the teweets in archive dataframe
        for tweet_id in df_twitter['tweet_id']:
            try:
                page = api.get_status(tweet_id, tweet_mode = 'extended')
                # Print one page to look at the structure of the returned file
                # and the names of attributes
                # print(json.dumps(page, indent = 4))
                #break

                favorites = page['favorite_count'] # How many favorites the tweet had
                retweets = page['retweet_count'] # Count of the retweet
                user_followers = page['user']['followers_count'] # How many followers the user h
                user_favourites = page['user']['favourites_count'] # How many favorites the user
                date_time = page['created_at'] # The date and time of the creation

                df_list.append({'tweet_id': int(tweet_id),
                                'favorites': int(favorites),
                                'retweets': int(retweets),
                                'user_followers': int(user_followers),
                                'user_favourites': int(user_favourites),
                                'date_time': pd.to_datetime(date_time)})

            # Catch the exceptions of the TweepError
```

16

```python
        except Exception as e:
            print(str(tweet_id)+ " _ " + str(e))
            error_list.append(tweet_id)

888202515573088257 _ [{'code': 144, 'message': 'No status found with that ID.'}]
873697596434513921 _ [{'code': 144, 'message': 'No status found with that ID.'}]
869988702071779329 _ [{'code': 144, 'message': 'No status found with that ID.'}]
866816280283807744 _ [{'code': 144, 'message': 'No status found with that ID.'}]
861769973181624320 _ [{'code': 144, 'message': 'No status found with that ID.'}]
845459076796616705 _ [{'code': 144, 'message': 'No status found with that ID.'}]
842892208864923648 _ [{'code': 144, 'message': 'No status found with that ID.'}]
837012587749474308 _ [{'code': 144, 'message': 'No status found with that ID.'}]
827228250799742977 _ [{'code': 144, 'message': 'No status found with that ID.'}]
802247111496568832 _ [{'code': 144, 'message': 'No status found with that ID.'}]
775096608509886464 _ [{'code': 144, 'message': 'No status found with that ID.'}]
Rate limit reached. Sleeping for: 735
Rate limit reached. Sleeping for: 734


In [7]: ee_list = []
        for e in error_list:
            try:
                favorites = page['favorite_count']
                retweets = page['retweet_count']
                user_followers = page['user']['followers_count']
                user_favourites = page['user']['favourites_count']
                date_time = page['created_at']

                df_list.append({'tweet_id': int(tweet_id),
                                'favorites': int(favorites),
                                'retweets': int(retweets),
                                'user_followers': int(user_followers),
                                'user_favourites': int(user_favourites),
                                'date_time': pd.to_datetime(date_time)})

            except Exception:
                print(str(tweet_id)+ " _ " + str(e))
                ee_list.append(e)
In [8]: json_tweets = pd.DataFrame(df_list, columns = ['tweet_id', 'favorites', 'retweets',
                                                       'user_followers', 'user_favourites', 'dat

        # Save the dataFrame in file
        json_tweets.to_csv('tweet_json.txt', encoding = 'utf-8', index=False)

In [6]: json_tweets = pd.read_csv('tweet_json.txt', encoding = 'utf-8')
        json_tweets

Out[6]:              tweet_id  favorites  retweets  user_followers  \
        0    892420643555336193      38690      8558         6992402
```

1	892177421306343426	33159	6291	6992402
2	891815181378084864	24962	4174	6992402
3	891689557279858688	42071	8690	6992402
4	891327558926688256	40227	9450	6992402
5	891087950875897856	20170	3128	6992402
6	890971913173991426	11814	2081	6992402
7	890729181411237888	65354	18979	6992402
8	890609185150312448	27724	4280	6992402
9	890240255349198849	31861	7446	6992402
10	890006608113172480	30582	7364	6992402
11	889880896479866881	27722	4990	6992403
12	889665388333682689	48011	10108	6992403
13	889638837579907072	27109	4565	6992403
14	889531135344209921	15058	2243	6992403
15	889278841981685760	25242	5448	6992403
16	888917238123831296	29002	4516	6992403
17	888804989199671297	25537	4362	6992403
18	888554962724278272	19854	3596	6992403
19	888078434458587136	21702	3509	6992403
20	887705289381826560	30109	5415	6992403
21	887517139158093824	46122	11717	6992403
22	887473957103951883	68942	18308	6992403
23	887343217045368832	33596	10448	6992403
24	887101392804085760	30453	5987	6992403
25	886983233522544640	35066	7808	6992403
26	886736880519319552	12039	3312	6992403
27	886680336477933568	22361	4488	6992403
28	886366144734445568	21146	3207	6992403
29	886267009285017600	116	4	6992403
...	...	...	...	...
2326	666268910803644416	104	35	6992426
2327	661104133288665088	14346	6635	6992426
2328	661102155909144576	80	13	6992426
2329	660099513787052032	156	68	6992426
2330	660094000022159362	164	74	6992426
2331	666082916733198337	119	45	6992426
2332	666073100786774016	322	164	6992426
2333	666071193221509120	148	62	6992426
2334	666063827256086533	476	220	6992426
2335	666058600524156928	112	57	6992426
2336	666057090499244032	298	142	6992426
2337	666055525042405380	434	252	6992426
2338	666051853826850816	1224	853	6992426
2339	666050758794694657	132	58	6992426
2340	666049248165822465	109	41	6992426
2341	666044226329800704	299	141	6992426
2342	666033412701032449	125	45	6992426
2343	666029285002620928	129	47	6992426

```
2344  666020888022790149       2563      516        6992426
2345  666020888022790149       2563      516        6992426
2346  666020888022790149       2563      516        6992426
2347  666020888022790149       2563      516        6992426
2348  666020888022790149       2563      516        6992426
2349  666020888022790149       2563      516        6992426
2350  666020888022790149       2563      516        6992426
2351  666020888022790149       2563      516        6992426
2352  666020888022790149       2563      516        6992426
2353  666020888022790149       2563      516        6992426
2354  666020888022790149       2563      516        6992426
2355  666020888022790149       2563      516        6992426

      user_favourites         date_time
0              134549  2017-08-01 16:23:56
1              134549  2017-08-01 00:17:27
2              134549  2017-07-31 00:18:03
3              134549  2017-07-30 15:58:51
4              134549  2017-07-29 16:00:24
5              134549  2017-07-29 00:08:17
6              134549  2017-07-28 16:27:12
7              134549  2017-07-28 00:22:40
8              134549  2017-07-27 16:25:51
9              134549  2017-07-26 15:59:51
10             134549  2017-07-26 00:31:25
11             134549  2017-07-25 16:11:53
12             134549  2017-07-25 01:55:32
13             134549  2017-07-25 00:10:02
14             134549  2017-07-24 17:02:04
15             134549  2017-07-24 00:19:32
16             134549  2017-07-23 00:22:39
17             134549  2017-07-22 16:56:37
18             134549  2017-07-22 00:23:06
19             134549  2017-07-20 16:49:33
20             134549  2017-07-19 16:06:48
21             134549  2017-07-19 03:39:09
22             134549  2017-07-19 00:47:34
23             134549  2017-07-18 16:08:03
24             134549  2017-07-18 00:07:08
25             134549  2017-07-17 16:17:36
26             134549  2017-07-16 23:58:41
27             134549  2017-07-16 20:14:00
28             134549  2017-07-15 23:25:31
29             134549  2017-07-15 16:51:35
...               ...                  ...
2326           134549  2015-11-16 14:57:41
2327           134549  2015-11-16 04:02:55
2328           134549  2015-11-16 03:55:04
```

```
2329          134549  2015-11-16 03:44:34
2330          134549  2015-11-16 03:22:39
2331          134549  2015-11-16 02:38:37
2332          134549  2015-11-16 01:59:36
2333          134549  2015-11-16 01:52:02
2334          134549  2015-11-16 01:22:45
2335          134549  2015-11-16 01:01:59
2336          134549  2015-11-16 00:55:59
2337          134549  2015-11-16 00:49:46
2338          134549  2015-11-16 00:35:11
2339          134549  2015-11-16 00:30:50
2340          134549  2015-11-16 00:24:50
2341          134549  2015-11-16 00:04:52
2342          134549  2015-11-15 23:21:54
2343          134549  2015-11-15 23:05:30
2344          134549  2015-11-15 22:32:08
2345          134549  2015-11-15 22:32:08
2346          134549  2015-11-15 22:32:08
2347          134549  2015-11-15 22:32:08
2348          134549  2015-11-15 22:32:08
2349          134549  2015-11-15 22:32:08
2350          134549  2015-11-15 22:32:08
2351          134549  2015-11-15 22:32:08
2352          134549  2015-11-15 22:32:08
2353          134549  2015-11-15 22:32:08
2354          134549  2015-11-15 22:32:08
2355          134549  2015-11-15 22:32:08

[2356 rows x 6 columns]
```

2 Access

```
In [12]: df_twitter.head()

Out[12]:              tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
         0  892420643555336193                    NaN                  NaN
         1  892177421306343426                    NaN                  NaN
         2  891815181378084864                    NaN                  NaN
         3  891689557279858688                    NaN                  NaN
         4  891327558926688256                    NaN                  NaN


                      timestamp  \
         0  2017-08-01 16:23:56 +0000
         1  2017-08-01 00:17:27 +0000
         2  2017-07-31 00:18:03 +0000
         3  2017-07-30 15:58:51 +0000
         4  2017-07-29 16:00:24 +0000
```

```
                                                      source  \
0  <a href="http://twitter.com/download/iphone" r...
1  <a href="http://twitter.com/download/iphone" r...
2  <a href="http://twitter.com/download/iphone" r...
3  <a href="http://twitter.com/download/iphone" r...
4  <a href="http://twitter.com/download/iphone" r...

                                                text  retweeted_status_id  \
0  This is Phineas. He's a mystical boy. Only eve...                  NaN
1  This is Tilly. She's just checking pup on you...                   NaN
2  This is Archie. He is a rare Norwegian Pouncin...                  NaN
3  This is Darla. She commenced a snooze mid meal...                  NaN
4  This is Franklin. He would like you to stop ca...                  NaN

   retweeted_status_user_id retweeted_status_timestamp  \
0                       NaN                        NaN
1                       NaN                        NaN
2                       NaN                        NaN
3                       NaN                        NaN
4                       NaN                        NaN

                                      expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643...                13
1  https://twitter.com/dog_rates/status/892177421...                13
2  https://twitter.com/dog_rates/status/891815181...                12
3  https://twitter.com/dog_rates/status/891689557...                13
4  https://twitter.com/dog_rates/status/891327558...                12

   rating_denominator      name doggo floofer pupper puppo
0                  10   Phineas  None    None   None  None
1                  10     Tilly  None    None   None  None
2                  10    Archie  None    None   None  None
3                  10     Darla  None    None   None  None
4                  10  Franklin  None    None   None  None

In [7]: df_twitter

Out[7]:                 tweet_id  in_reply_to_status_id  in_reply_to_user_id  \
        0     892420643555336193                    NaN                  NaN
        1     892177421306343426                    NaN                  NaN
        2     891815181378084864                    NaN                  NaN
        3     891689557279858688                    NaN                  NaN
        4     891327558926688256                    NaN                  NaN
        5     891087950875897856                    NaN                  NaN
        6     890971913173991426                    NaN                  NaN
        7     890729181411237888                    NaN                  NaN
        8     890609185150312448                    NaN                  NaN
```

9	890240255349198849	NaN	NaN
10	890006608113172480	NaN	NaN
11	889880896479866881	NaN	NaN
12	889665388333682689	NaN	NaN
13	889638837579907072	NaN	NaN
14	889531135344209921	NaN	NaN
15	889278841981685760	NaN	NaN
16	888917238123831296	NaN	NaN
17	888804989199671297	NaN	NaN
18	888554962724278272	NaN	NaN
19	888202515573088257	NaN	NaN
20	888078434458587136	NaN	NaN
21	887705289381826560	NaN	NaN
22	887517139158093824	NaN	NaN
23	887473957103951883	NaN	NaN
24	887343217045368832	NaN	NaN
25	887101392804085760	NaN	NaN
26	886983233522544640	NaN	NaN
27	886736880519319552	NaN	NaN
28	886680336477933568	NaN	NaN
29	886366144734445568	NaN	NaN
...	...	...	...
2326	666411507551481857	NaN	NaN
2327	666407126856765440	NaN	NaN
2328	666396247373291520	NaN	NaN
2329	666373753744588802	NaN	NaN
2330	666362758909284353	NaN	NaN
2331	666353288456101888	NaN	NaN
2332	666345417576210432	NaN	NaN
2333	666337882303524864	NaN	NaN
2334	666293911632134144	NaN	NaN
2335	666287406224695296	NaN	NaN
2336	666273097616637952	NaN	NaN
2337	666268910803644416	NaN	NaN
2338	666104133288665088	NaN	NaN
2339	666102155909144576	NaN	NaN
2340	666099513787052032	NaN	NaN
2341	666094000022159362	NaN	NaN
2342	666082916733198337	NaN	NaN
2343	666073100786774016	NaN	NaN
2344	666071193221509120	NaN	NaN
2345	666063827256086533	NaN	NaN
2346	666058600524156928	NaN	NaN
2347	666057090499244032	NaN	NaN
2348	666055525042405380	NaN	NaN
2349	666051853826850816	NaN	NaN
2350	666050758794694657	NaN	NaN
2351	666049248165822465	NaN	NaN

```
2352   666044226329800704                                   NaN                  NaN
2353   666033412701032449                                   NaN                  NaN
2354   666029285002620928                                   NaN                  NaN
2355   666020888022790149                                   NaN                  NaN

                           timestamp  \
0        2017-08-01 16:23:56 +0000
1        2017-08-01 00:17:27 +0000
2        2017-07-31 00:18:03 +0000
3        2017-07-30 15:58:51 +0000
4        2017-07-29 16:00:24 +0000
5        2017-07-29 00:08:17 +0000
6        2017-07-28 16:27:12 +0000
7        2017-07-28 00:22:40 +0000
8        2017-07-27 16:25:51 +0000
9        2017-07-26 15:59:51 +0000
10       2017-07-26 00:31:25 +0000
11       2017-07-25 16:11:53 +0000
12       2017-07-25 01:55:32 +0000
13       2017-07-25 00:10:02 +0000
14       2017-07-24 17:02:04 +0000
15       2017-07-24 00:19:32 +0000
16       2017-07-23 00:22:39 +0000
17       2017-07-22 16:56:37 +0000
18       2017-07-22 00:23:06 +0000
19       2017-07-21 01:02:36 +0000
20       2017-07-20 16:49:33 +0000
21       2017-07-19 16:06:48 +0000
22       2017-07-19 03:39:09 +0000
23       2017-07-19 00:47:34 +0000
24       2017-07-18 16:08:03 +0000
25       2017-07-18 00:07:08 +0000
26       2017-07-17 16:17:36 +0000
27       2017-07-16 23:58:41 +0000
28       2017-07-16 20:14:00 +0000
29       2017-07-15 23:25:31 +0000
...                            ...
2326     2015-11-17 00:24:19 +0000
2327     2015-11-17 00:06:54 +0000
2328     2015-11-16 23:23:41 +0000
2329     2015-11-16 21:54:18 +0000
2330     2015-11-16 21:10:36 +0000
2331     2015-11-16 20:32:58 +0000
2332     2015-11-16 20:01:42 +0000
2333     2015-11-16 19:31:45 +0000
2334     2015-11-16 16:37:02 +0000
2335     2015-11-16 16:11:11 +0000
2336     2015-11-16 15:14:19 +0000
```

```
2337    2015-11-16 14:57:41 +0000
2338    2015-11-16 04:02:55 +0000
2339    2015-11-16 03:55:04 +0000
2340    2015-11-16 03:44:34 +0000
2341    2015-11-16 03:22:39 +0000
2342    2015-11-16 02:38:37 +0000
2343    2015-11-16 01:59:36 +0000
2344    2015-11-16 01:52:02 +0000
2345    2015-11-16 01:22:45 +0000
2346    2015-11-16 01:01:59 +0000
2347    2015-11-16 00:55:59 +0000
2348    2015-11-16 00:49:46 +0000
2349    2015-11-16 00:35:11 +0000
2350    2015-11-16 00:30:50 +0000
2351    2015-11-16 00:24:50 +0000
2352    2015-11-16 00:04:52 +0000
2353    2015-11-15 23:21:54 +0000
2354    2015-11-15 23:05:30 +0000
2355    2015-11-15 22:32:08 +0000

                                                         source  \
0      <a href="http://twitter.com/download/iphone" r...
1      <a href="http://twitter.com/download/iphone" r...
2      <a href="http://twitter.com/download/iphone" r...
3      <a href="http://twitter.com/download/iphone" r...
4      <a href="http://twitter.com/download/iphone" r...
5      <a href="http://twitter.com/download/iphone" r...
6      <a href="http://twitter.com/download/iphone" r...
7      <a href="http://twitter.com/download/iphone" r...
8      <a href="http://twitter.com/download/iphone" r...
9      <a href="http://twitter.com/download/iphone" r...
10     <a href="http://twitter.com/download/iphone" r...
11     <a href="http://twitter.com/download/iphone" r...
12     <a href="http://twitter.com/download/iphone" r...
13     <a href="http://twitter.com/download/iphone" r...
14     <a href="http://twitter.com/download/iphone" r...
15     <a href="http://twitter.com/download/iphone" r...
16     <a href="http://twitter.com/download/iphone" r...
17     <a href="http://twitter.com/download/iphone" r...
18     <a href="http://twitter.com/download/iphone" r...
19     <a href="http://twitter.com/download/iphone" r...
20     <a href="http://twitter.com/download/iphone" r...
21     <a href="http://twitter.com/download/iphone" r...
22     <a href="http://twitter.com/download/iphone" r...
23     <a href="http://twitter.com/download/iphone" r...
24     <a href="http://twitter.com/download/iphone" r...
25     <a href="http://twitter.com/download/iphone" r...
26     <a href="http://twitter.com/download/iphone" r...
```

```
27       <a href="http://twitter.com/download/iphone" r...
28       <a href="http://twitter.com/download/iphone" r...
29       <a href="http://twitter.com/download/iphone" r...
...                                                    ...
2326     <a href="http://twitter.com/download/iphone" r...
2327     <a href="http://twitter.com/download/iphone" r...
2328     <a href="http://twitter.com/download/iphone" r...
2329     <a href="http://twitter.com/download/iphone" r...
2330     <a href="http://twitter.com/download/iphone" r...
2331     <a href="http://twitter.com/download/iphone" r...
2332     <a href="http://twitter.com/download/iphone" r...
2333     <a href="http://twitter.com/download/iphone" r...
2334     <a href="http://twitter.com/download/iphone" r...
2335     <a href="http://twitter.com/download/iphone" r...
2336     <a href="http://twitter.com/download/iphone" r...
2337     <a href="http://twitter.com/download/iphone" r...
2338     <a href="http://twitter.com/download/iphone" r...
2339     <a href="http://twitter.com/download/iphone" r...
2340     <a href="http://twitter.com/download/iphone" r...
2341     <a href="http://twitter.com/download/iphone" r...
2342     <a href="http://twitter.com/download/iphone" r...
2343     <a href="http://twitter.com/download/iphone" r...
2344     <a href="http://twitter.com/download/iphone" r...
2345     <a href="http://twitter.com/download/iphone" r...
2346     <a href="http://twitter.com/download/iphone" r...
2347     <a href="http://twitter.com/download/iphone" r...
2348     <a href="http://twitter.com/download/iphone" r...
2349     <a href="http://twitter.com/download/iphone" r...
2350     <a href="http://twitter.com/download/iphone" r...
2351     <a href="http://twitter.com/download/iphone" r...
2352     <a href="http://twitter.com/download/iphone" r...
2353     <a href="http://twitter.com/download/iphone" r...
2354     <a href="http://twitter.com/download/iphone" r...
2355     <a href="http://twitter.com/download/iphone" r...


                                                    text   retweeted_status_id  \
0       This is Phineas. He's a mystical boy. Only eve...                  NaN
1       This is Tilly. She's just checking pup on you...                   NaN
2       This is Archie. He is a rare Norwegian Pouncin...                  NaN
3       This is Darla. She commenced a snooze mid meal...                  NaN
4       This is Franklin. He would like you to stop ca...                  NaN
5       Here we have a majestic great white breaching ...                  NaN
6       Meet Jax. He enjoys ice cream so much he gets ...                  NaN
7       When you watch your owner call another dog a g...                   NaN
8       This is Zoey. She doesn't want to be one of th...                  NaN
9       This is Cassie. She is a college pup. Studying...                  NaN
10      This is Koda. He is a South Australian decksha...                  NaN
11      This is Bruno. He is a service shark. Only get...                  NaN
```

```
12      Here's a puppo that seems to be on the fence a...              NaN
13      This is Ted. He does his best. Sometimes that'...              NaN
14      This is Stuart. He's sporting his favorite fan...              NaN
15      This is Oliver. You're witnessing one of his m...              NaN
16      This is Jim. He found a fren. Taught him how t...              NaN
17      This is Zeke. He has a new stick. Very proud o...              NaN
18      This is Ralphus. He's powering up. Attempting ...              NaN
19      RT @dog_rates: This is Canela. She attempted s...     8.874740e+17
20      This is Gerald. He was just told he didn't get...              NaN
21      This is Jeffrey. He has a monopoly on the pool...              NaN
22      I've yet to rate a Venezuelan Hover Wiener. Th...              NaN
23      This is Canela. She attempted some fancy porch...              NaN
24      You may not have known you needed to see this ...              NaN
25      This... is a Jubilant Antarctic House Bear. We...              NaN
26      This is Maya. She's very shy. Rarely leaves he...              NaN
27      This is Mingus. He's a wonderful father to his...              NaN
28      This is Derek. He's late for a dog meeting. 13...              NaN
29      This is Roscoe. Another pupper fallen victim t...              NaN
...                                      ...                           ...
2326    This is quite the dog. Gets really excited whe...              NaN
2327    This is a southern Vesuvius bumblegruff. Can d...              NaN
2328    Oh goodness. A super rare northeast Qdoba kang...              NaN
2329    Those are sunglasses and a jean jacket. 11/10 ...              NaN
2330    Unique dog here. Very small. Lives in containe...              NaN
2331    Here we have a mixed Asiago from the Galápagos...              NaN
2332    Look at this jokester thinking seat belt laws ...              NaN
2333    This is an extremely rare horned Parthenon. No...              NaN
2334    This is a funny dog. Weird toes. Won't come do...              NaN
2335    This is an Albanian 3 1/2 legged  Episcopalian...              NaN
2336        Can take selfies 11/10 https://t.co/ws2AMaNwPW           NaN
2337    Very concerned about fellow dog trapped in com...              NaN
2338    Not familiar with this breed. No tail (weird)...              NaN
2339    Oh my. Here you are seeing an Adobe Setter giv...              NaN
2340    Can stand on stump for what seems like a while...              NaN
2341    This appears to be a Mongolian Presbyterian mi...              NaN
2342    Here we have a well-established sunblockerspan...              NaN
2343    Let's hope this flight isn't Malaysian (lol). ...              NaN
2344    Here we have a northern speckled Rhododendron...              NaN
2345    This is the happiest dog you will ever see. Ve...              NaN
2346    Here is the Rand Paul of retrievers folks! He'...              NaN
2347    My oh my. This is a rare blond Canadian terrie...              NaN
2348    Here is a Siberian heavily armored polar bear ...              NaN
2349    This is an odd dog. Hard on the outside but lo...              NaN
2350    This is a truly beautiful English Wilson Staff...              NaN
2351    Here we have a 1949 1st generation vulpix. Enj...              NaN
2352    This is a purebred Piers Morgan. Loves to Netf...              NaN
2353    Here is a very happy pup. Big fan of well-main...              NaN
2354    This is a western brown Mitsubishi terrier. Up...              NaN
```

```
2355  Here we have a Japanese Irish Setter. Lost eye...                         NaN

      retweeted_status_user_id retweeted_status_timestamp  \
0                          NaN                        NaN
1                          NaN                        NaN
2                          NaN                        NaN
3                          NaN                        NaN
4                          NaN                        NaN
5                          NaN                        NaN
6                          NaN                        NaN
7                          NaN                        NaN
8                          NaN                        NaN
9                          NaN                        NaN
10                         NaN                        NaN
11                         NaN                        NaN
12                         NaN                        NaN
13                         NaN                        NaN
14                         NaN                        NaN
15                         NaN                        NaN
16                         NaN                        NaN
17                         NaN                        NaN
18                         NaN                        NaN
19                4.196984e+09   2017-07-19 00:47:34 +0000
20                         NaN                        NaN
21                         NaN                        NaN
22                         NaN                        NaN
23                         NaN                        NaN
24                         NaN                        NaN
25                         NaN                        NaN
26                         NaN                        NaN
27                         NaN                        NaN
28                         NaN                        NaN
29                         NaN                        NaN
...                        ...                        ...
2326                       NaN                        NaN
2327                       NaN                        NaN
2328                       NaN                        NaN
2329                       NaN                        NaN
2330                       NaN                        NaN
2331                       NaN                        NaN
2332                       NaN                        NaN
2333                       NaN                        NaN
2334                       NaN                        NaN
2335                       NaN                        NaN
2336                       NaN                        NaN
2337                       NaN                        NaN
2338                       NaN                        NaN
2339                       NaN                        NaN
```

```
2340                    NaN                    NaN
2341                    NaN                    NaN
2342                    NaN                    NaN
2343                    NaN                    NaN
2344                    NaN                    NaN
2345                    NaN                    NaN
2346                    NaN                    NaN
2347                    NaN                    NaN
2348                    NaN                    NaN
2349                    NaN                    NaN
2350                    NaN                    NaN
2351                    NaN                    NaN
2352                    NaN                    NaN
2353                    NaN                    NaN
2354                    NaN                    NaN
2355                    NaN                    NaN


                                          expanded_urls  rating_numerator  \
0      https://twitter.com/dog_rates/status/892420643...                13
1      https://twitter.com/dog_rates/status/892177421...                13
2      https://twitter.com/dog_rates/status/891815181...                12
3      https://twitter.com/dog_rates/status/891689557...                13
4      https://twitter.com/dog_rates/status/891327558...                12
5      https://twitter.com/dog_rates/status/891087950...                13
6      https://gofundme.com/ydvmve-surgery-for-jax,ht...                13
7      https://twitter.com/dog_rates/status/890729181...                13
8      https://twitter.com/dog_rates/status/890609185...                13
9      https://twitter.com/dog_rates/status/890240255...                14
10     https://twitter.com/dog_rates/status/890006608...                13
11     https://twitter.com/dog_rates/status/889880896...                13
12     https://twitter.com/dog_rates/status/889665388...                13
13     https://twitter.com/dog_rates/status/889638837...                12
14     https://twitter.com/dog_rates/status/889531135...                13
15     https://twitter.com/dog_rates/status/889278841...                13
16     https://twitter.com/dog_rates/status/888917238...                12
17     https://twitter.com/dog_rates/status/888804989...                13
18     https://twitter.com/dog_rates/status/888554962...                13
19     https://twitter.com/dog_rates/status/887473957...                13
20     https://twitter.com/dog_rates/status/888078434...                12
21     https://twitter.com/dog_rates/status/887705289...                13
22     https://twitter.com/dog_rates/status/887517139...                14
23     https://twitter.com/dog_rates/status/887473957...                13
24     https://twitter.com/dog_rates/status/887343217...                13
25     https://twitter.com/dog_rates/status/887101392...                12
26     https://twitter.com/dog_rates/status/886983233...                13
27     https://www.gofundme.com/mingusneedsus,https:/...                13
28     https://twitter.com/dog_rates/status/886680336...                13
29     https://twitter.com/dog_rates/status/886366144...                12
```

```
      ...                                                       ...            ...
2326  https://twitter.com/dog_rates/status/666411507...          2
2327  https://twitter.com/dog_rates/status/666407126...          7
2328  https://twitter.com/dog_rates/status/666396247...          9
2329  https://twitter.com/dog_rates/status/666373753...         11
2330  https://twitter.com/dog_rates/status/666362758...          6
2331  https://twitter.com/dog_rates/status/666353288...          8
2332  https://twitter.com/dog_rates/status/666345417...         10
2333  https://twitter.com/dog_rates/status/666337882...          9
2334  https://twitter.com/dog_rates/status/666293911...          3
2335  https://twitter.com/dog_rates/status/666287406...          1
2336  https://twitter.com/dog_rates/status/666273097...         11
2337  https://twitter.com/dog_rates/status/666268910...         10
2338  https://twitter.com/dog_rates/status/666104133...          1
2339  https://twitter.com/dog_rates/status/666102155...         11
2340  https://twitter.com/dog_rates/status/666099513...          8
2341  https://twitter.com/dog_rates/status/666094000...          9
2342  https://twitter.com/dog_rates/status/666082916...          6
2343  https://twitter.com/dog_rates/status/666073100...         10
2344  https://twitter.com/dog_rates/status/666071193...          9
2345  https://twitter.com/dog_rates/status/666063827...         10
2346  https://twitter.com/dog_rates/status/666058600...          8
2347  https://twitter.com/dog_rates/status/666057090...          9
2348  https://twitter.com/dog_rates/status/666055525...         10
2349  https://twitter.com/dog_rates/status/666051853...          2
2350  https://twitter.com/dog_rates/status/666050758...         10
2351  https://twitter.com/dog_rates/status/666049248...          5
2352  https://twitter.com/dog_rates/status/666044226...          6
2353  https://twitter.com/dog_rates/status/666033412...          9
2354  https://twitter.com/dog_rates/status/666029285...          7
2355  https://twitter.com/dog_rates/status/666020888...          8

      rating_denominator      name   doggo floofer  pupper  puppo
0                     10   Phineas    None    None    None   None
1                     10     Tilly    None    None    None   None
2                     10    Archie    None    None    None   None
3                     10     Darla    None    None    None   None
4                     10  Franklin    None    None    None   None
5                     10      None    None    None    None   None
6                     10       Jax    None    None    None   None
7                     10      None    None    None    None   None
8                     10      Zoey    None    None    None   None
9                     10    Cassie   doggo    None    None   None
10                    10      Koda    None    None    None   None
11                    10     Bruno    None    None    None   None
12                    10      None    None    None    None  puppo
13                    10       Ted    None    None    None   None
14                    10    Stuart    None    None    None  puppo
```

29

15	10	Oliver	None	None	None	None
16	10	Jim	None	None	None	None
17	10	Zeke	None	None	None	None
18	10	Ralphus	None	None	None	None
19	10	Canela	None	None	None	None
20	10	Gerald	None	None	None	None
21	10	Jeffrey	None	None	None	None
22	10	such	None	None	None	None
23	10	Canela	None	None	None	None
24	10	None	None	None	None	None
25	10	None	None	None	None	None
26	10	Maya	None	None	None	None
27	10	Mingus	None	None	None	None
28	10	Derek	None	None	None	None
29	10	Roscoe	None	None	pupper	None
...	...	...	...	...	...	...
2326	10	quite	None	None	None	None
2327	10	a	None	None	None	None
2328	10	None	None	None	None	None
2329	10	None	None	None	None	None
2330	10	None	None	None	None	None
2331	10	None	None	None	None	None
2332	10	None	None	None	None	None
2333	10	an	None	None	None	None
2334	10	a	None	None	None	None
2335	2	an	None	None	None	None
2336	10	None	None	None	None	None
2337	10	None	None	None	None	None
2338	10	None	None	None	None	None
2339	10	None	None	None	None	None
2340	10	None	None	None	None	None
2341	10	None	None	None	None	None
2342	10	None	None	None	None	None
2343	10	None	None	None	None	None
2344	10	None	None	None	None	None
2345	10	the	None	None	None	None
2346	10	the	None	None	None	None
2347	10	a	None	None	None	None
2348	10	a	None	None	None	None
2349	10	an	None	None	None	None
2350	10	a	None	None	None	None
2351	10	None	None	None	None	None
2352	10	a	None	None	None	None
2353	10	a	None	None	None	None
2354	10	a	None	None	None	None
2355	10	None	None	None	None	None

[2356 rows x 17 columns]

```
In [14]: df_twitter.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 17 columns):
tweet_id                    2356 non-null int64
in_reply_to_status_id       78 non-null float64
in_reply_to_user_id         78 non-null float64
timestamp                   2356 non-null object
source                      2356 non-null object
text                        2356 non-null object
retweeted_status_id         181 non-null float64
retweeted_status_user_id    181 non-null float64
retweeted_status_timestamp  181 non-null object
expanded_urls               2297 non-null object
rating_numerator            2356 non-null int64
rating_denominator          2356 non-null int64
name                        2356 non-null object
doggo                       2356 non-null object
floofer                     2356 non-null object
pupper                      2356 non-null object
puppo                       2356 non-null object
dtypes: float64(4), int64(3), object(10)
memory usage: 313.0+ KB


In [15]: df_twitter.describe()
```

Out[15]:

	tweet_id	in_reply_to_status_id	in_reply_to_user_id	\
count	2.356000e+03	7.800000e+01	7.800000e+01	
mean	7.427716e+17	7.455079e+17	2.014171e+16	
std	6.856705e+16	7.582492e+16	1.252797e+17	
min	6.660209e+17	6.658147e+17	1.185634e+07	
25%	6.783989e+17	6.757419e+17	3.086374e+08	
50%	7.196279e+17	7.038708e+17	4.196984e+09	
75%	7.993373e+17	8.257804e+17	4.196984e+09	
max	8.924206e+17	8.862664e+17	8.405479e+17	

	retweeted_status_id	retweeted_status_user_id	rating_numerator	\
count	1.810000e+02	1.810000e+02	2356.000000	
mean	7.720400e+17	1.241698e+16	13.126486	
std	6.236928e+16	9.599254e+16	45.876648	
min	6.661041e+17	7.832140e+05	0.000000	
25%	7.186315e+17	4.196984e+09	10.000000	
50%	7.804657e+17	4.196984e+09	11.000000	
75%	8.203146e+17	4.196984e+09	12.000000	
max	8.874740e+17	7.874618e+17	1776.000000	

```
          rating_denominator
count            2356.000000
mean               10.455433
std                 6.745237
min                 0.000000
25%                10.000000
50%                10.000000
75%                10.000000
max               170.000000
```

In [16]: df_image.head()

Out[16]:   tweet_id   jpg_url \
         0 666020888022790149 https://pbs.twimg.com/media/CT4udn0WwAAOaMy.jpg
         1 666029285002620928 https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg
         2 666033412701032449 https://pbs.twimg.com/media/CT4521TWwAEvMyu.jpg
         3 666044226329800704 https://pbs.twimg.com/media/CT5Dr8HUEAA-lEu.jpg
         4 666049248165822465 https://pbs.twimg.com/media/CT5IQmsXIAAKY4A.jpg

            img_num   p1   p1_conf p1_dog   p2 \
         0   1 Welsh_springer_spaniel 0.465074   True   collie
         1   1   redbone 0.506826   True miniature_pinscher
         2   1   German_shepherd 0.596461   True   malinois
         3   1   Rhodesian_ridgeback 0.408143   True   redbone
         4   1   miniature_pinscher 0.560311   True   Rottweiler

            p2_conf p2_dog   p3   p3_conf p3_dog
         0 0.156665   True   Shetland_sheepdog 0.061428   True
         1 0.074192   True   Rhodesian_ridgeback 0.072010   True
         2 0.138584   True   bloodhound 0.116197   True
         3 0.360687   True   miniature_pinscher 0.222752   True
         4 0.243682   True   Doberman 0.154629   True
```

In [17]: df_image.tail()

Out[17]:                 tweet_id                                          jpg_url  \
         2070  891327558926688256  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg
         2071  891689557279858688  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg
         2072  891815181378084864  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg
         2073  892177421306343426  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg
         2074  892420643555336193  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg


               img_num            p1   p1_conf  p1_dog                   p2   p2_conf  \
         2070        2        basset  0.555712    True     English_springer  0.225770
         2071        1   paper_towel  0.170278   False   Labrador_retriever  0.168086
         2072        1      Chihuahua  0.716012    True             malamute  0.078253
         2073        1      Chihuahua  0.323581    True             Pekinese  0.090647
         2074        1        orange  0.097049   False                bagel  0.085851
```

```
             p2_dog                               p3     p3_conf   p3_dog
2070           True   German_short-haired_pointer  0.175219    True
2071           True                       spatula  0.040836   False
2072           True                        kelpie  0.031379    True
2073           True                      papillon  0.068957    True
2074          False                        banana  0.076110   False
```

In [18]: df_image.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2075 entries, 0 to 2074
Data columns (total 12 columns):
tweet_id    2075 non-null int64
jpg_url     2075 non-null object
img_num     2075 non-null int64
p1          2075 non-null object
p1_conf     2075 non-null float64
p1_dog      2075 non-null bool
p2          2075 non-null object
p2_conf     2075 non-null float64
p2_dog      2075 non-null bool
p3          2075 non-null object
p3_conf     2075 non-null float64
p3_dog      2075 non-null bool
dtypes: bool(3), float64(3), int64(2), object(4)
memory usage: 152.1+ KB
```

In [19]: df_image.describe()

```
Out[19]:           tweet_id       img_num       p1_conf        p2_conf        p3_conf
         count  2.075000e+03   2075.000000   2075.000000   2.075000e+03   2.075000e+03
         mean   7.384514e+17      1.203855      0.594548   1.345886e-01   6.032417e-02
         std    6.785203e+16      0.561875      0.271174   1.006657e-01   5.090593e-02
         min    6.660209e+17      1.000000      0.044333   1.011300e-08   1.740170e-10
         25%    6.764835e+17      1.000000      0.364412   5.388625e-02   1.622240e-02
         50%    7.119988e+17      1.000000      0.588230   1.181810e-01   4.944380e-02
         75%    7.932034e+17      1.000000      0.843855   1.955655e-01   9.180755e-02
         max    8.924206e+17      4.000000      1.000000   4.880140e-01   2.734190e-01
```

In [11]: json_tweets.head()

```
Out[11]:            tweet_id  favorites  retweets  user_followers  user_favourites  \
         0  892420643555336193      38690      8558         6992402           134549
         1  892177421306343426      33159      6291         6992402           134549
         2  891815181378084864      24962      4174         6992402           134549
         3  891689557279858688      42071      8690         6992402           134549
         4  891327558926688256      40227      9450         6992402           134549
```

```
               date_time
      0   2017-08-01 16:23:56
      1   2017-08-01 00:17:27
      2   2017-07-31 00:18:03
      3   2017-07-30 15:58:51
      4   2017-07-29 16:00:24


In [12]: json_tweets.tail()

Out[12]:                   tweet_id  favorites  retweets  user_followers  \
      2351  666020888022790149       2563       516         6992426
      2352  666020888022790149       2563       516         6992426
      2353  666020888022790149       2563       516         6992426
      2354  666020888022790149       2563       516         6992426
      2355  666020888022790149       2563       516         6992426


              user_favourites             date_time
      2351            134549   2015-11-15 22:32:08
      2352            134549   2015-11-15 22:32:08
      2353            134549   2015-11-15 22:32:08
      2354            134549   2015-11-15 22:32:08
      2355            134549   2015-11-15 22:32:08


In [13]: json_tweets.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2356 entries, 0 to 2355
Data columns (total 6 columns):
tweet_id          2356 non-null int64
favorites         2356 non-null int64
retweets          2356 non-null int64
user_followers    2356 non-null int64
user_favourites   2356 non-null int64
date_time         2356 non-null object
dtypes: int64(5), object(1)
memory usage: 110.5+ KB



In [14]: json_tweets.describe()

Out[14]:            tweet_id      favorites       retweets  user_followers  \
      count  2.356000e+03    2356.000000    2356.000000    2.356000e+03
      mean   7.419379e+17    8018.380730    3003.025891    6.992430e+06
      std    6.837471e+16   12081.792098    5005.958273    1.140488e+02
      min    6.660209e+17       0.000000       0.000000    6.992402e+06
      25%    6.779127e+17    1404.000000     596.750000    6.992410e+06
      50%    7.183508e+17    3512.500000    1399.000000    6.992419e+06
      75%    7.986836e+17    9911.250000    3500.500000    6.992425e+06
      max    8.924206e+17  142980.000000   77127.000000    6.993395e+06
```

34

```
           user_favourites
    count      2356.000000
    mean     134549.014007
    std           0.117543
    min      134549.000000
    25%      134549.000000
    50%      134549.000000
    75%      134549.000000
    max      134550.000000
```

3 Clean

3.0.1 Quality

df_twitter

- in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id should be integers and not float
- In several columns null objects are not null
- Name column have invalid names
- We want originals tweets having ratings not retweets

df_image

- Missing values from image dataset
- Some tweet_ids have the same jpg_url
- Some tweets are have 2 different tweet_id one redirect to the other

json_tweets

- duplicated records

3.0.2 Tideness

- Unnecessary details
- Various stages of dogs in columns instead of rows archives dataset

```
In [7]: df_master = pd.merge(df_twitter,df_image)

In [8]: df_master = pd.merge(df_master, json_tweets)

In [7]: df_master.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2080 entries, 0 to 2079
Data columns (total 33 columns):
tweet_id                    2080 non-null int64
in_reply_to_status_id         23 non-null float64
```

```
in_reply_to_user_id           23 non-null float64
timestamp                     2080 non-null object
source                        2080 non-null object
text                          2080 non-null object
retweeted_status_id           75 non-null float64
retweeted_status_user_id      75 non-null float64
retweeted_status_timestamp    75 non-null object
expanded_urls                 2080 non-null object
rating_numerator              2080 non-null int64
rating_denominator            2080 non-null int64
name                          2080 non-null object
doggo                         2080 non-null object
floofer                       2080 non-null object
pupper                        2080 non-null object
puppo                         2080 non-null object
jpg_url                       2080 non-null object
img_num                       2080 non-null int64
p1                            2080 non-null object
p1_conf                       2080 non-null float64
p1_dog                        2080 non-null bool
p2                            2080 non-null object
p2_conf                       2080 non-null float64
p2_dog                        2080 non-null bool
p3                            2080 non-null object
p3_conf                       2080 non-null float64
p3_dog                        2080 non-null bool
favorites                     2080 non-null int64
retweets                      2080 non-null int64
user_followers                2080 non-null int64
user_favourites               2080 non-null int64
date_time                     2080 non-null object
dtypes: bool(3), float64(7), int64(8), object(15)
memory usage: 509.8+ KB


In [9]: df_master = df_master[pd.isnull(df_master.retweeted_status_id)]
        df_master = df_master.drop_duplicates()
        df_master = df_master.dropna(subset = ['jpg_url'])

In [10]: df_master = df_master.drop('retweeted_status_id', 1)

In [11]: df_master = df_master.drop('retweeted_status_user_id', 1)
         df_master = df_master.drop('retweeted_status_timestamp', 1)

In [12]: df_master = df_master.drop('date_time', 1)

In [12]: df_master.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 0 to 2068
```

```
Data columns (total 29 columns):
tweet_id                1994 non-null int64
in_reply_to_status_id   23 non-null float64
in_reply_to_user_id     23 non-null float64
timestamp               1994 non-null object
source                  1994 non-null object
text                    1994 non-null object
expanded_urls           1994 non-null object
rating_numerator        1994 non-null int64
rating_denominator      1994 non-null int64
name                    1994 non-null object
doggo                   1994 non-null object
floofer                 1994 non-null object
pupper                  1994 non-null object
puppo                   1994 non-null object
jpg_url                 1994 non-null object
img_num                 1994 non-null int64
p1                      1994 non-null object
p1_conf                 1994 non-null float64
p1_dog                  1994 non-null bool
p2                      1994 non-null object
p2_conf                 1994 non-null float64
p2_dog                  1994 non-null bool
p3                      1994 non-null object
p3_conf                 1994 non-null float64
p3_dog                  1994 non-null bool
favorites               1994 non-null int64
retweets                1994 non-null int64
user_followers          1994 non-null int64
user_favourites         1994 non-null int64
dtypes: bool(3), float64(5), int64(8), object(13)
memory usage: 426.5+ KB


In [13]: print(df_master.doggo.value_counts())
         print(df_master.floofer.value_counts())
         print(df_master.pupper.value_counts())
         print(df_master.puppo.value_counts())

None      1920
doggo       74
Name: doggo, dtype: int64
None       1986
floofer       8
Name: floofer, dtype: int64
None       1782
pupper      212
Name: pupper, dtype: int64
```

```
None         1971
puppo          23
Name: puppo, dtype: int64
```

```python
In [14]: melt_columns = ['doggo', 'floofer', 'pupper', 'puppo']
         not_to_melt_columns = [ x for x in df_master.columns.tolist() if x not in melt_columns]
         df_master = pd.melt(df_master, id_vars = not_to_melt_columns, value_vars = melt_columns
                             var_name = 'stages', value_name = 'dog_stage')
```

```python
In [15]: df_master = df_master.drop('stages', 1)
```

```python
In [14]: print(df_master.dog_stage.value_counts())
```

```
None         7659
pupper        212
doggo          74
puppo          23
floofer         8
Name: dog_stage, dtype: int64
```

```python
In [16]: df_master = df_master.sort_values('dog_stage').drop_duplicates('tweet_id', keep = 'last
```

```python
In [16]: print(df_master.dog_stage.value_counts())
```

```
None         1688
pupper        212
doggo          63
puppo          23
floofer         8
Name: dog_stage, dtype: int64
```

```python
In [17]: df_master.apply(lambda x : len(x.unique()))
```

```
Out[17]: tweet_id                 1994
         in_reply_to_status_id      23
         in_reply_to_user_id         2
         timestamp                1994
         source                      3
         text                     1994
         expanded_urls            1994
         rating_numerator           34
         rating_denominator         15
         name                      936
         jpg_url                  1994
         img_num                     4
         p1                        374
```

```
            p1_conf                  1991
            p1_dog                      2
            p2                        401
            p2_conf                  1989
            p2_dog                      2
            p3                        405
            p3_conf                  1991
            p3_dog                      2
            favorites                1837
            retweets                 1600
            user_followers             33
            user_favourites            2
            dog_stage                   5
            dtype: int64

In [39]: df_master.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1994 entries, 1918 to 6764
Data columns (total 26 columns):
tweet_id              1994 non-null int64
in_reply_to_status_id   23 non-null float64
in_reply_to_user_id     23 non-null float64
timestamp             1994 non-null object
source                1994 non-null object
text                  1994 non-null object
expanded_urls         1994 non-null object
rating_numerator      1994 non-null int64
rating_denominator    1994 non-null int64
name                  1994 non-null object
jpg_url               1994 non-null object
img_num               1994 non-null int64
p1                    1994 non-null object
p1_conf               1994 non-null float64
p1_dog                1994 non-null bool
p2                    1994 non-null object
p2_conf               1994 non-null float64
p2_dog                1994 non-null bool
p3                    1994 non-null object
p3_conf               1994 non-null float64
p3_dog                1994 non-null bool
favorites             1994 non-null int64
retweets              1994 non-null int64
user_followers        1994 non-null int64
user_favourites       1994 non-null int64
dog_stage             1994 non-null object
dtypes: bool(3), float64(5), int64(8), object(10)
memory usage: 379.7+ KB
```

```
Out[39]: 134549    1967
         134550      27
         Name: user_favourites, dtype: int64

In [18]: df_master['in_reply_to_user_id'].value_counts()
         df_master['source'].value_counts()
         df_master['user_favourites'].value_counts()

Out[18]: 134549    1967
         134550      27
         Name: user_favourites, dtype: int64

In [19]: df_master = df_master.drop(['in_reply_to_status_id', 'in_reply_to_user_id', 'user_favou

In [20]: df_master['source'] = df_master['source'].apply(lambda x: re.findall(r'>(.*)<', x)[0])

In [21]: df_master

Out[21]:              tweet_id                    timestamp              source  \
         1918  667405339315146752  2015-11-19 18:13:27 +0000  Twitter for iPhone
         1917  667435689202614272  2015-11-19 20:14:03 +0000  Twitter for iPhone
         1916  667437278097252352  2015-11-19 20:20:22 +0000  Twitter for iPhone
         1915  667443425659232256  2015-11-19 20:44:47 +0000  Twitter for iPhone
         1914  667453023279554560  2015-11-19 21:22:56 +0000  Twitter Web Client
         1913  667455448082227200  2015-11-19 21:32:34 +0000  Twitter Web Client
         1912  667470559035432960  2015-11-19 22:32:36 +0000  Twitter Web Client
         1911  667491009379606528  2015-11-19 23:53:52 +0000  Twitter Web Client
         1910  667495797102141441  2015-11-20 00:12:54 +0000  Twitter Web Client
         1909  667502640335572993  2015-11-20 00:40:05 +0000  Twitter Web Client
         1908  667509364010450944  2015-11-20 01:06:48 +0000  Twitter Web Client
         1907  667517642048163840  2015-11-20 01:39:42 +0000  Twitter Web Client
         1906  667524857454854144  2015-11-20 02:08:22 +0000  Twitter Web Client
         1905  667530908589760512  2015-11-20 02:32:25 +0000  Twitter Web Client
         1904  667534815156183040  2015-11-20 02:47:56 +0000  Twitter Web Client
         1903  667538891197542400  2015-11-20 03:04:08 +0000  Twitter Web Client
         1902  667544320556335104  2015-11-20 03:25:43 +0000  Twitter Web Client
         1901  667546741521195010  2015-11-20 03:35:20 +0000  Twitter Web Client
         1900  667549055577362432  2015-11-20 03:44:31 +0000  Twitter Web Client
         1899  667724302356258817  2015-11-20 15:20:54 +0000  Twitter Web Client
         1898  667728196545200128  2015-11-20 15:36:22 +0000  Twitter Web Client
         1897  667766675769573376  2015-11-20 18:09:16 +0000  Twitter Web Client
         1896  667773195014021121  2015-11-20 18:35:10 +0000  Twitter Web Client
         1895  667782464991965184  2015-11-20 19:12:01 +0000  Twitter for iPhone
         1894  667793409583771648  2015-11-20 19:55:30 +0000  Twitter for iPhone
         1893  667801013445750784  2015-11-20 20:25:43 +0000  Twitter for iPhone
         1892  667806454573760512  2015-11-20 20:47:20 +0000  Twitter for iPhone
         1919  667393430834667520  2015-11-19 17:26:08 +0000  Twitter for iPhone
         1891  667832474953625600  2015-11-20 22:30:44 +0000  Twitter for iPhone
         1920  667369227918143488  2015-11-19 15:49:57 +0000  Twitter for iPhone
```

```
     ...                ...                   ...                  ...
5200  690015576308211712   2016-01-21 03:38:27 +0000   Twitter for iPhone
5513  675113801096802304   2015-12-11 00:44:07 +0000   Twitter for iPhone
5204  689905486972461056   2016-01-20 20:21:00 +0000   Twitter for iPhone
4064  874296783580663808   2017-06-12 16:06:11 +0000   Twitter for iPhone
5543  674447403907457024   2015-12-09 04:36:06 +0000   Twitter for iPhone
5057  702598099714314240   2016-02-24 20:56:55 +0000   Twitter for iPhone
5209  689623661272240129   2016-01-20 01:41:08 +0000   Twitter for iPhone
6281  825026590719483904   2017-01-27 17:04:02 +0000   Twitter for iPhone
5996  889531135344209921   2017-07-24 17:02:04 +0000   Twitter for iPhone
6680  752519690950500352   2016-07-11 15:07:30 +0000   Twitter for iPhone
6130  855851453814013952   2017-04-22 18:31:02 +0000   Twitter for iPhone
6291  822872901745569793   2017-01-21 18:26:02 +0000   Twitter for iPhone
6691  751132876104687617   2016-07-07 19:16:47 +0000   Twitter for iPhone
6448  793195938047070209   2016-10-31 21:00:23 +0000   Twitter for iPhone
6060  874012996292530176   2017-06-11 21:18:31 +0000   Twitter for iPhone
5994  889665388333682689   2017-07-25 01:55:32 +0000   Twitter for iPhone
6116  859607811541651456   2017-05-03 03:17:27 +0000   Twitter for iPhone
6481  787717603741622272   2016-10-16 18:11:26 +0000   Twitter for iPhone
6308  819952236453363712   2017-01-13 17:00:21 +0000   Twitter for iPhone
6795  738537504001953792   2016-06-03 01:07:16 +0000   Twitter for iPhone
6753  744995568523612160   2016-06-20 20:49:19 +0000   Twitter for iPhone
6661  756275833623502848   2016-07-21 23:53:04 +0000   Twitter for iPhone
6516  780931614150983680   2016-09-28 00:46:20 +0000   Twitter for iPhone
6402  802239329049477120   2016-11-25 19:55:35 +0000   Twitter for iPhone
6394  803773340896923648   2016-11-30 01:31:12 +0000   Twitter for iPhone
6463  790946055508652032   2016-10-25 16:00:09 +0000   Twitter for iPhone
6087  867421006826221569   2017-05-24 16:44:18 +0000   Twitter for iPhone
6044  878776093423087618   2017-06-25 00:45:22 +0000   Twitter for iPhone
6279  825535076884762624   2017-01-29 02:44:34 +0000   Twitter for iPhone
6764  743253157753532416   2016-06-16 01:25:36 +0000   Twitter for iPhone

                                                             text  \
1918   This is Biden. Biden just tripped... 7/10 http...
1917              Ermergerd 12/10 https://t.co/PQni2sjPsm
1916   Never seen this breed before. Very pointy pup...
1915   Exotic dog here. Long neck. Weird paws. Obsess...
1914   Meet Cupcake. I would do unspeakable things fo...
1913   This is Reese and Twips. Reese protects Twips...
1912   This is a northern Wahoo named Kohl. He runs t...
1911   Two dogs in this one. Both are rare Jujitsu Py...
1910   This is Philippe from Soviet Russia. Commandin...
1909   Say hello to Hall and Oates. Oates is winking ...
1908   This a Norwegian Pewterschmidt named Tickles. ...
1907   This is Dook &amp; Milo. Dook is struggling to...
1906   Another topnotch dog. His name is Big Jumpy Ra...
1905   Meet Naphaniel. He doesn't necessarily enjoy h...
1904   This is Frank (pronounced "Fronq"). Too many b...
```

```
1903  This is a southwest Coriander named Klint. Hat...
1902  This is Kial. Kial is either wearing a cape, w...
1901  Here is George. George took a selfie of his ne...
1900  Never seen dog like this. Breathes heavy. Tilt...
1899  What a dog to start the day with. Very calm. L...
1898  Meet Olive. He comes to spot by tree to remini...
1897  This is Calvin. He is a Luxembourgian Mayo. Ha...
1896  This is a rare Hungarian Pinot named Jessiga. ...
1895  Super rare dog. Endangered (?). Thinks it's fu...
1894  Dogs only please. Small cows and other non can...
1893                OMIGOD 12/10 https://t.co/SVMF4Frf1w
1892  This is Filup. He is overcome with joy after f...
1919  This is Fwed. He is a Canadian Asian Taylormad...
1891  THE EYES 12/10\n\nI'm sorry. These are suppose...
1920  Here we have a neat pup. Very white. Cool shad...
...                                                    ...
5200  This pupper can only sleep on shoes. It's a cr...
5513  Meet Zuzu. He just graduated college. Astute p...
5204  Say hello to Gizmo. He's quite the pupper. Con...
4064  This is Jed. He may be the fanciest pupper in ...
5543  This pupper just wants a belly rub. This puppe...
5057  This is Sansa. She's gotten too big for her ch...
5209  This is Lucy. She's terrified of the stuffed b...
6281  Say hello to Pablo. He's one gorgeous puppo. A...
5996  This is Stuart. He's sporting his favorite fan...
6680  Hopefully this puppo on a swing will help get ...
6130  Here's a puppo participating in the #ScienceMa...
6291  Here's a super supportive puppo participating ...
6691  This is Cooper. He's just so damn happy. 10/10...
6448  Say hello to Lily. She's pupset that her costu...
6060  This is Sebastian. He can't see all the colors...
5994  Here's a puppo that seems to be on the fence a...
6116  Sorry for the lack of posts today. I came home...
6481  This is Tonks. She is a service puppo. Can hea...
6308  This is Oliver. He has dreams of being a servi...
6795  This is Bayley. She fell asleep trying to esca...
6753  This is Abby. She got her face stuck in a glas...
6661  When ur older siblings get to play in the deep...
6516  I want to finally rate this iconic puppo who t...
6402  This is Loki. He'll do your taxes for you. Can...
6394  This is Diogi. He fell in the pool as soon as ...
6463  This is Betty. She's assisting with the dishes...
6087  This is Shikha. She just watched you drop a sk...
6044  This is Snoopy. He's a proud #PrideMonthPuppo...
6279  Here's a very loving and accepting puppo. Appe...
6764  This is Kilo. He cannot reach the snackum. Nif...

                              expanded_urls  rating_numerator  \
```

```
1918    https://twitter.com/dog_rates/status/667405339...            7
1917    https://twitter.com/dog_rates/status/667435689...           12
1916    https://twitter.com/dog_rates/status/667437278...           10
1915    https://twitter.com/dog_rates/status/667443425...            6
1914    https://twitter.com/dog_rates/status/667453023...           11
1913    https://twitter.com/dog_rates/status/667455448...            7
1912    https://twitter.com/dog_rates/status/667470559...           11
1911    https://twitter.com/dog_rates/status/667491009...            7
1910    https://twitter.com/dog_rates/status/667495797...            9
1909    https://twitter.com/dog_rates/status/667502640...           11
1908    https://twitter.com/dog_rates/status/667509364...           12
1907    https://twitter.com/dog_rates/status/667517642...            8
1906    https://twitter.com/dog_rates/status/667524857...           12
1905    https://twitter.com/dog_rates/status/667530908...           10
1904    https://twitter.com/dog_rates/status/667534815...            8
1903    https://twitter.com/dog_rates/status/667538891...            9
1902    https://twitter.com/dog_rates/status/667544320...           10
1901    https://twitter.com/dog_rates/status/667546741...            9
1900    https://twitter.com/dog_rates/status/667549055...            1
1899    https://twitter.com/dog_rates/status/667724302...            7
1898    https://twitter.com/dog_rates/status/667728196...           11
1897    https://twitter.com/dog_rates/status/667766675...            9
1896    https://twitter.com/dog_rates/status/667773195...            8
1895    https://twitter.com/dog_rates/status/667782464...            9
1894    https://twitter.com/dog_rates/status/667793409...            8
1893    https://twitter.com/dog_rates/status/667801013...           12
1892    https://twitter.com/dog_rates/status/667806454...           10
1919    https://twitter.com/dog_rates/status/667393430...            8
1891    https://twitter.com/dog_rates/status/667832474...           12
1920    https://twitter.com/dog_rates/status/667369227...           10
...                                                      ...        ...
5200    https://twitter.com/dog_rates/status/690015576...           12
5513    https://twitter.com/dog_rates/status/675113801...           10
5204    https://twitter.com/dog_rates/status/689905486...           11
4064    https://twitter.com/dog_rates/status/874296783...           13
5543    https://twitter.com/dog_rates/status/674447403...           10
5057    https://twitter.com/dog_rates/status/702598099...           11
5209    https://twitter.com/dog_rates/status/689623661...           10
6281    https://www.gofundme.com/my-puppys-double-cata...           12
5996    https://twitter.com/dog_rates/status/889531135...           13
6680    https://twitter.com/dog_rates/status/752519690...           11
6130    https://twitter.com/dog_rates/status/855851453...           13
6291    https://twitter.com/dog_rates/status/822872901...           13
6691    https://twitter.com/dog_rates/status/751132876...           10
6448    https://twitter.com/dog_rates/status/793195938...           12
6060    https://twitter.com/dog_rates/status/874012996...           13
5994    https://twitter.com/dog_rates/status/889665388...           13
6116    https://twitter.com/dog_rates/status/859607811...           13
```

```
6481   https://twitter.com/dog_rates/status/787717603...          13
6308   https://www.gofundme.com/servicedogoliver,http...          13
6795   https://twitter.com/dog_rates/status/738537504...          11
6753   https://twitter.com/dog_rates/status/744995568...           9
6661   https://twitter.com/dog_rates/status/756275833...          10
6516   https://twitter.com/dog_rates/status/780931614...          13
6402   https://twitter.com/dog_rates/status/802239329...          12
6394   https://twitter.com/dog_rates/status/803773340...          12
6463   https://twitter.com/dog_rates/status/790946055...          12
6087   https://twitter.com/dog_rates/status/867421006...          12
6044   https://twitter.com/dog_rates/status/878776093...          13
6279   https://twitter.com/dog_rates/status/825535076...          14
6764   https://twitter.com/dog_rates/status/743253157...          10

       rating_denominator        name  \
1918                   10       Biden
1917                   10        None
1916                   10        None
1915                   10        None
1914                   10     Cupcake
1913                   10       Reese
1912                   10           a
1911                   10        None
1910                   10     Philippe
1909                   10        Hall
1908                   10        None
1907                   10        Dook
1906                   10        None
1905                   10    Naphaniel
1904                   10       Frank
1903                   10           a
1902                   10        Kial
1901                   10      George
1900                   10        None
1899                   10        None
1898                   10       Olive
1897                   10      Calvin
1896                   10           a
1895                   10        None
1894                   10        None
1893                   10        None
1892                   10       Filup
1919                   10        Fwed
1891                   10        None
1920                   10        None
...                   ...        ...
5200                   10        None
5513                   10        Zuzu
```

```
5204              10        Gizmo
4064              10          Jed
5543              10         None
5057              10        Sansa
5209              10         Lucy
6281              10        Pablo
5996              10       Stuart
6680              10         None
6130              10         None
6291              10         None
6691              10       Cooper
6448              10         Lily
6060              10    Sebastian
5994              10         None
6116              10         None
6481              10        Tonks
6308              10       Oliver
6795              10       Bayley
6753              10         Abby
6661              10         None
6516              10         None
6402              10         Loki
6394              10        Diogi
6463              10        Betty
6087              10       Shikha
6044              10       Snoopy
6279              10         None
6764              10         Kilo

                                             jpg_url  img_num    ...        \
1918  https://pbs.twimg.com/media/CUMZnmhUEAEbtis.jpg        1    ...
1917  https://pbs.twimg.com/media/CUM1OHCW4AEgGSi.jpg        1    ...
1916  https://pbs.twimg.com/media/CUM2qWaWoAUZ06L.jpg        1    ...
1915  https://pbs.twimg.com/media/CUM8QZwW4AAVsBl.jpg        1    ...
1914  https://pbs.twimg.com/media/CUNE_OSUwAAdHhX.jpg        1    ...
1913  https://pbs.twimg.com/media/CUNHMXTU8AAS3HH.jpg        1    ...
1912  https://pbs.twimg.com/media/CUNU78YWEAECmpB.jpg        1    ...
1911  https://pbs.twimg.com/media/CUNniSlUYAEj1Jl.jpg        1    ...
1910  https://pbs.twimg.com/media/CUNr4-7UwAAg2lq.jpg        1    ...
1909  https://pbs.twimg.com/media/CUNyHTMUYAAQVch.jpg        1    ...
1908  https://pbs.twimg.com/media/CUN4Or5UAAAa5K4.jpg        1    ...
1907  https://pbs.twimg.com/media/CUN_wiBUkAAakTO.jpg        1    ...
1906  https://pbs.twimg.com/media/CUOGUfJW4AA_eni.jpg        1    ...
1905  https://pbs.twimg.com/media/CUOL0uGUkAAx7yh.jpg        1    ...
1904  https://pbs.twimg.com/media/CUOPYI5UcAAj_nO.jpg        1    ...
1903  https://pbs.twimg.com/media/CUOTFZOW4AABsfW.jpg        1    ...
1902  https://pbs.twimg.com/media/CUOYBbbWIAAXQGU.jpg        1    ...
1901  https://pbs.twimg.com/media/CUOaOWXWcAAO_Jy.jpg        1    ...
```

```
1900   https://pbs.twimg.com/media/CUOcVCwWsAERUKY.jpg              1      ...
1899   https://pbs.twimg.com/media/CUQ7tv3W4AA3KlI.jpg              1      ...
1898   https://pbs.twimg.com/media/CUQ_QahUAAAVQjn.jpg              1      ...
1897   https://pbs.twimg.com/media/CURiQMnUAAAPT2M.jpg              1      ...
1896   https://pbs.twimg.com/media/CURoLrOVEAAaWdR.jpg              1      ...
1895   https://pbs.twimg.com/media/CURwm3cUkAARcO6.jpg              1      ...
1894   https://pbs.twimg.com/media/CUR6jqVWsAEgGot.jpg              1      ...
1893   https://pbs.twimg.com/media/CUSBemVUEAAn-6V.jpg              1      ...
1892   https://pbs.twimg.com/media/CUSGbXeVAAAgztZ.jpg              1      ...
1919   https://pbs.twimg.com/media/CUMOyd3XIAAl13H.jpg              1      ...
1891   https://pbs.twimg.com/media/CUSeGFNW4AAyyHC.jpg              1      ...
1920   https://pbs.twimg.com/media/CUL4xR9UkAEdlJ6.jpg              1      ...
...                                                          ...     ...    ...
5200   https://pbs.twimg.com/media/CZNtgWhWkAAbq3W.jpg              2      ...
5513   https://pbs.twimg.com/media/CV58a4nXAAApywo.jpg              1      ...
5204   https://pbs.twimg.com/media/CZMJYCRVAAE35Wk.jpg              4      ...
4064   https://pbs.twimg.com/media/DCIgSR0XgAANEOY.jpg              1      ...
5543   https://pbs.twimg.com/media/CVweVUfW4AACPwI.jpg              1      ...
5057   https://pbs.twimg.com/media/CcAhPevW8AAoknv.jpg              1      ...
5209   https://pbs.twimg.com/media/CZIJD2SWIAMJgNI.jpg              1      ...
6281   https://pbs.twimg.com/media/C3MVTeHWcAAGNfx.jpg              2      ...
5996   https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg              1      ...
6680   https://pbs.twimg.com/media/CnF8qVDWYAAhOg1.jpg              3      ...
6130   https://pbs.twimg.com/media/C-CYWrvWAAU8AXH.jpg              1      ...
6291   https://pbs.twimg.com/media/C2tugXLXgAArJO4.jpg              1      ...
6691   https://pbs.twimg.com/media/CmyPXNOW8AEtaJ-.jpg              1      ...
6448   https://pbs.twimg.com/media/CwH_foYWgAEvTyI.jpg              2      ...
6060   https://pbs.twimg.com/media/DCEeLxjXsAAvNSM.jpg              2      ...
5994   https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg              1      ...
6116   https://pbs.twimg.com/media/C-3wvtxXcAUTuBE.jpg              1      ...
6481   https://pbs.twimg.com/media/Cu6I9vvWIAAZG0a.jpg              3      ...
6308   https://pbs.twimg.com/media/C2EONHNWQAUWxkP.jpg              1      ...
6795   https://pbs.twimg.com/media/Cj_P7rSUgAAYQbz.jpg              1      ...
6753   https://pbs.twimg.com/media/ClbBg4WWEAMjwJu.jpg              1      ...
6661   https://pbs.twimg.com/media/Cn7U2xlW8AI9Pqp.jpg              1      ...
6516   https://pbs.twimg.com/media/CtZtJxAXEAAyPGd.jpg              1      ...
6402   https://pbs.twimg.com/media/CyIgaTEVEAA-9zS.jpg              2      ...
6394   https://pbs.twimg.com/media/CyeTku-XcAALkBd.jpg              2      ...
6463   https://pbs.twimg.com/media/CvoBPWRWgAA4het.jpg              1      ...
6087   https://pbs.twimg.com/media/DAmyy8FXYAIH8Ty.jpg              1      ...
6044   https://pbs.twimg.com/media/DDIKMXzW0AEibje.jpg              2      ...
6279   https://pbs.twimg.com/media/C3TjvitXAAAI-QH.jpg              1      ...
6764   https://pbs.twimg.com/media/ClCQzFUUYAA5vAu.jpg              1      ...


                            p2    p2_conf  p2_dog                      p3  \
1918               Leonberg  0.127998    True          golden_retriever
1917     miniature_pinscher  0.000450    True   black-and-tan_coonhound
1916              bath_towel  0.006300   False                    badger
```

46

1915	drake	0.006918	False	hen
1914	French_bulldog	0.056639	True	Staffordshire_bullterrier
1913	Irish_terrier	0.054933	True	Yorkshire_terrier
1912	pug	0.223427	True	Lakeland_terrier
1911	ice_bear	0.132264	False	weasel
1910	Christmas_stocking	0.118651	False	ski_mask
1909	golden_retriever	0.001688	True	beagle
1908	Labrador_retriever	0.119256	True	golden_retriever
1907	standard_poodle	0.084572	True	cocker_spaniel
1906	dhole	0.092435	False	Chesapeake_Bay_retriever
1905	kuvasz	0.146391	True	Labrador_retriever
1904	Cardigan	0.307407	True	cocker_spaniel
1903	silky_terrier	0.300313	True	Australian_terrier
1902	Pembroke	0.312958	True	Chihuahua
1901	miniature_poodle	0.202225	True	teddy
1900	spotlight	0.007737	False	lampshade
1899	bighorn	0.125119	False	ram
1898	golden_retriever	0.293744	True	Labrador_retriever
1897	tow_truck	0.074734	False	jeep
1896	pug	0.093494	True	ice_bear
1895	hummingbird	0.083011	False	African_grey
1894	English_setter	0.451219	True	Great_Dane
1893	Chesapeake_Bay_retriever	0.262239	True	curly-coated_retriever
1892	Chihuahua	0.187155	True	Brabancon_griffon
1919	Border_collie	0.271963	True	collie
1891	bath_towel	0.146789	False	Chihuahua
1920	bath_towel	0.127285	False	Christmas_stocking
...	...	...	...	...
5200	Siberian_husky	0.033084	True	Eskimo_dog
5513	quill	0.108807	False	joystick
5204	Shetland_sheepdog	0.023675	True	chow
4064	miniature_poodle	0.277191	True	toy_poodle
5543	malinois	0.244649	True	bull_mastiff
5057	badger	0.133584	False	Siamese_cat
5209	mashed_potato	0.208564	False	Labrador_retriever
6281	Siberian_husky	0.467678	True	malamute
5996	Labrador_retriever	0.013834	True	redbone
6680	Labrador_retriever	0.000010	True	Eskimo_dog
6130	Labrador_retriever	0.115138	True	groenendael
6291	Labrador_retriever	0.160329	True	Irish_terrier
6691	Chesapeake_Bay_retriever	0.038254	True	golden_retriever
6448	golden_retriever	0.074100	True	Chihuahua
6060	Pembroke	0.116622	True	kelpie
5994	Cardigan	0.027356	True	basenji
6116	Irish_setter	0.024099	True	Labrador_retriever
6481	malinois	0.004920	True	kelpie
6308	Staffordshire_bullterrier	0.036221	True	Italian_greyhound
6795	gibbon	0.028942	False	Pembroke

6753	Shih-Tzu	0.146336	True		Tibetan_terrier
6661	Irish_terrier	0.086981	True		bloodhound
6516	necklace	0.065462	False		chain
6402	Siberian_husky	0.335774	True		malamute
6394	redbone	0.059707	True		Irish_terrier
6463	golden_retriever	0.245773	True		chow
6087	Siberian_husky	0.381330	True		malamute
6044	whippet	0.150487	True		Ibizan_hound
6279	Tibetan_mastiff	0.147940	True	black-and-tan_coonhound
6764	Siberian_husky	0.368137	True		Eskimo_dog

	p3_conf	p3_dog	favorites	retweets	user_followers	dog_stage
1918	0.069357	True	474	224	6992426	None
1917	0.000157	True	310	85	6992426	None
1916	0.000966	False	458	248	6992426	None
1915	0.005255	False	803	604	6992426	None
1914	0.054018	True	320	93	6992426	None
1913	0.040576	True	195	65	6992426	None
1912	0.073316	True	264	100	6992426	None
1911	0.005730	False	547	235	6992426	None
1910	0.092482	False	540	286	6992426	None
1909	0.000712	True	544	227	6992426	None
1908	0.082549	True	6956	2210	6992426	None
1907	0.081347	True	377	198	6992426	None
1906	0.088122	True	1745	1162	6992426	None
1905	0.046184	True	487	253	6992426	None
1904	0.033158	True	841	559	6992426	None
1903	0.053412	True	209	69	6992426	None
1902	0.071960	True	889	547	6992426	None
1901	0.004047	False	341	130	6992426	None
1900	0.001901	False	5961	2386	6993394	None
1899	0.074673	False	502	332	6992426	None
1898	0.270673	True	383	156	6992426	None
1897	0.012773	False	456	228	6992426	None
1896	0.069038	False	240	59	6992426	None
1895	0.054247	False	423	253	6992426	None
1894	0.008164	True	718	341	6992426	None
1893	0.048920	True	338	98	6992426	None
1892	0.112799	True	1084	520	6992426	None
1919	0.073473	True	199	58	6992426	None
1891	0.104152	True	294	65	6992426	None
1920	0.028568	False	375	168	6992426	None
...	...	...	...	...	...	...
5200	0.016663	True	2681	807	6992420	pupper
5513	0.043312	False	2046	839	6992426	pupper
5204	0.007165	True	2593	759	6992421	pupper
4064	0.157402	True	26058	4150	6992403	pupper
5543	0.074820	True	1103	381	6992426	pupper

48

```
5057  0.074440   False    11044     3583    6992419    pupper
5209  0.077481    True     2391      715    6992421    pupper
6281  0.004976    True     6856     1432    6992407     puppo
5996  0.007958    True    15058     2243    6992403     puppo
6680  0.000001    True     7939     3790    6992416     puppo
6130  0.096100    True    46856    18650    6992403     puppo
6291  0.069126    True   142980    48957    6992407     puppo
6691  0.007610    True     5468     1438    6992416     puppo
6448  0.042339    True    16672     6326    6992411     puppo
6060  0.049182    True    34680    10574    6992403     puppo
5994  0.004633    True    48011    10108    6992403     puppo
6116  0.019285    True    19036     1640    6992403     puppo
6481  0.000853    True    11130     3125    6992410     puppo
6308  0.020412    True     5761     1322    6992407     puppo
6795  0.026498    True     5434     1708    6992418     puppo
6753  0.134269    True     3194      692    6992417     puppo
6661  0.086276    True     6941     1697    6992416     puppo
6516  0.036469   False    23535     8237    6993378     puppo
6402  0.134589    True     9877     2927    6992412     puppo
6394  0.034195    True    10954     3108    6992411     puppo
6463  0.039012    True    18122     5291    6992411     puppo
6087  0.001670    True    16341     2583    6992403     puppo
6044  0.039725    True    19371     4155    6992403     puppo
6279  0.024525    True    56057    19124    6992407     puppo
6764  0.177822    True     4516     1326    6992417     puppo

[1994 rows x 23 columns]
```

```python
In [22]: df_master.rating_numerator.value_counts()
         df_master.rating_denominator.value_counts()
         print(df_master[df_master.rating_denominator == 170]['text'][2842])
         print(df_master[df_master.rating_numerator == 1776]['text'][2720])
         print(df_master[df_master.tweet_id == 786709082849828864]['text'][2497])
         print(df_master['text'][1918])
         print(df_master['text'][1917])
         print(df_master['text'][1916])
         print(df_master['text'][1911])
```

```
Say hello to this unbelievably well behaved squad of doggos. 204/170 would try to pet all at onc
This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkh
This is Logan, the Chow who lived. He solemnly swears he's up to lots of good. H*ckin magical af
This is Biden. Biden just tripped... 7/10 https://t.co/3Fm9PwLju1
Ermergerd 12/10 https://t.co/PQni2sjPsm
Never seen this breed before. Very pointy pup. Hurts when you cuddle. Still cute tho. 10/10 http
Two dogs in this one. Both are rare Jujitsu Pythagoreans. One slightly whiter than other. Long l
```

```python
In [38]: ratings = df_master['text'].apply(lambda x: re.findall(r'(\d+(\.\d+)|(\d+))\/(\d+0)', x
```

```
In [42]: rating_numerator = []
         rating_denominator = []
         dogs_count = []

In [48]: for rate in ratings:

             if len(rate) == 0:
                 rating_numerator.append('NaN')
                 rating_denominator.append('NaN')
                 dogs_count.append(1)


             elif len(rate) == 1:
                 rating_numerator.append((float(rate[0][0]) / (float(rate[0][-1])/10)))
                 rating_denominator.append(float(rate[0][-1]))
                 dogs_count.append(float(rate[0][-1])/ 10)

             elif len(rate) > 1 and rate[0][-1] == '10':
                 rating_plus = 0
                 rating_avg = 0
                 for i in range(len(rate)):
                     rating_plus = rating_plus + float(rate[i][0])
                 result_avg = (rating_plus / len(rate))
                 rating_numerator.append(result_avg)
                 rating_denominator.append(10)
                 dogs_count.append(len(rate))
             else:
                 rating_numerator.append('Error')
                 rating_denominator.append('Error')
                 dogs_count.append('Error')

         df_master['new_rating_numerator'] = rating_numerator
         df_master['new_rating_denominator'] = rating_denominator
         df_master['dogs_count'] = dogs_count
         df_master['new_rating_numerator'].value_counts()

Out[48]: 12.0          453
         10.0          411
         11.0          399
         13.0          261
         9.0           152
         8.0            94
         7.0            52
         14.0           36
         6.0            32
         5.0            30
         3.0            19
         4.0            14
```

```
            2.0                      9
            1.0                      4
            8.5                      4
            9.5                      3
            7.5                      3
            10.5                     2
            Error                    2
            0.0                      2
            1776.0                   1
            9.75                     1
            4.5                      1
            5.5                      1
            6.5                      1
            NaN                      1
            11.26                    1
            13.5                     1
            9.666666666666666        1
            420.0                    1
            11.5                     1
            11.27                    1
            Name: new_rating_numerator, dtype: int64
```

In [49]: `print(df_master[df_master.new_rating_numerator == 'Error']['text'][2919])`
         `print(df_master[df_master.new_rating_numerator == 'Error']['text'][2885])`
         `print(df_master[df_master.new_rating_numerator == 1776.0]['text'][2720])`
         `print(df_master[df_master.new_rating_numerator == 420.0]['text'][3712])`

```
This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. Amazed af. 11/10 ht
Happy 4/20 from the squad! 13/10 for all https://t.co/eV1diwds8a
This is Atticus. He's quite simply America af. 1776/10 https://t.co/GRXwMxLBkh
After so many requests... here you go.

Good dogg. 420/10 https://t.co/yfAAo1gdeY
```

In [50]: `tweet_id_11 = df_master[df_master.new_rating_numerator == 'Error']['tweet_id'][2919]`
         `tweet_id_13 = df_master[df_master.new_rating_numerator == 'Error']['tweet_id'][2885]`

         `df_master.loc[df_master['tweet_id'] == tweet_id_11, 'new_rating_numerator'] = 11`
         `df_master.loc[df_master['tweet_id'] == tweet_id_13, 'new_rating_numerator'] = 13`

         `df_master.loc[df_master['dogs_count'] == 'Error', 'dogs_count'] = 1`
         `df_master.loc[df_master['new_rating_denominator'] == 'Error', 'new_rating_denominator']`

In [51]: `print(df_master.new_rating_numerator[df_master.tweet_id == tweet_id_11])`
         `print(df_master.new_rating_numerator[df_master.tweet_id == tweet_id_13])`

```
2919      11
Name: new_rating_numerator, dtype: object
```

```
2885       13
Name: new_rating_numerator, dtype: object


In [52]: df_master = df_master.drop(['rating_numerator', 'rating_denominator'], 1)

         # Rename columns
         df_master.rename(columns = {'new_rating_numerator': 'rating_numerator',
                                     'new_rating_denominator': 'rating_denominator'}, inplace =


         list(df_master)

Out[52]: ['tweet_id',
          'timestamp',
          'source',
          'text',
          'expanded_urls',
          'name',
          'jpg_url',
          'img_num',
          'p1',
          'p1_conf',
          'p1_dog',
          'p2',
          'p2_conf',
          'p2_dog',
          'p3',
          'p3_conf',
          'p3_dog',
          'favorites',
          'retweets',
          'user_followers',
          'dog_stage',
          'rating_numerator',
          'rating_denominator',
          'dogs_count']

In [53]: dog_names = []

         for text in df_master['text']:

             if text.startswith('This is ') and re.match(r'[A-Z].*', text.split()[2]):
                 dog_names.append(text.split()[2].strip(',').strip('.'))

             elif text.startswith('Meet ') and re.match(r'[A-Z].*', text.split()[1]):
                 dog_names.append(text.split()[1].strip(',').strip('.'))
```

```python
            elif text.startswith('Say hello to ') and re.match(r'[A-Z].*', text.split()[3]):
                dog_names.append(text.split()[3].strip(',').strip('.'))

            elif text.startswith('Here we have ') and re.match(r'[A-Z].*', text.split()[3]):
                dog_names.append(text.split()[3].strip(',').strip('.'))

            elif 'named' in text and re.match(r'[A-Z].*', text.split()[text.split().index('name
                dog_names.append(text.split()[text.split().index('named') + 1].strip(',').strip

            else:
                dog_names.append('NaN')

        # Test
        len(dog_names)

        df_master['dog_name'] = dog_names

        print("New column dog_name count \n", df_master.dog_name.value_counts())
        print("Old column name count \n", df_master.name.value_counts())
```

```
New column dog_name count
 NaN        625
Charlie      11
Cooper       10
Lucy         10
Oliver       10
Penny         9
Tucker        9
Winston       8
Sadie         8
Daisy         7
Lola          7
Jax           6
Toby          6
Bella         6
Stanley       6
Koda          6
Bo            6
Chester       5
Scout         5
Rusty         5
Louis         5
Milo          5
Oscar         5
Buddy         5
Bailey        5
Leo           5
Larry         4
```

```
Jack          4
Clarence      4
Archie        4
              ...
Karma         1
Ben           1
Gerbald       1
Derby         1
Glenn         1
Beemo         1
Ralphson      1
Jay           1
Herb          1
Damon         1
Pluto         1
Scott         1
Aubie         1
Jockson       1
Bruiser       1
Canela        1
Cermet        1
Ronnie        1
Kreggory      1
Marlee        1
Jacob         1
Reptar        1
Iggy          1
Tayzie        1
Pancho        1
Sully         1
Aja           1
Franq         1
Alf           1
Huxley        1
Name: dog_name, Length: 934, dtype: int64
Old column name count
 None            546
a                55
Charlie          11
Cooper           10
Lucy             10
Oliver           10
Penny             9
Tucker            9
Sadie             8
Winston           8
Toby              7
the               7
```

Lola	7
Daisy	7
Koda	6
an	6
Bella	6
Jax	6
Stanley	6
Bo	6
Buddy	5
Bailey	5
Louis	5
Oscar	5
Leo	5
Milo	5
Rusty	5
Chester	5
Scout	5
Dave	5
...	
Alexanderson	1
Ace	1
Bert	1
Dido	1
Rose	1
Strider	1
Aubie	1
Craig	1
Jockson	1
Bruiser	1
Canela	1
Cermet	1
Marlee	1
Reptar	1
Iggy	1
Tayzie	1
Sully	1
Aja	1
Franq	1
Alf	1
Jay	1
Beemo	1
Rodney	1
Glenn	1
Tater	1
Kial	1
Eve	1
Lenox	1
Glacier	1

```
Huxley                1
Name: name, Length: 936, dtype: int64


In [54]: df_master[df_master.dog_name == 'NaN']

Out[54]:                  tweet_id                  timestamp                 source  \
         1917  667435689202614272  2015-11-19 20:14:03 +0000  Twitter for iPhone
         1916  667437278097252352  2015-11-19 20:20:22 +0000  Twitter for iPhone
         1915  667443425659232256  2015-11-19 20:44:47 +0000  Twitter for iPhone
         1911  667491009379606528  2015-11-19 23:53:52 +0000  Twitter Web Client
         1906  667524857454854144  2015-11-20 02:08:22 +0000  Twitter Web Client
         1901  667546741521195010  2015-11-20 03:35:20 +0000  Twitter Web Client
         1900  667549055577362432  2015-11-20 03:44:31 +0000  Twitter Web Client
         1899  667724302356258817  2015-11-20 15:20:54 +0000  Twitter Web Client
         1895  667782464991965184  2015-11-20 19:12:01 +0000  Twitter for iPhone
         1894  667793409583771648  2015-11-20 19:55:30 +0000  Twitter for iPhone
         1893  667801013445750784  2015-11-20 20:25:43 +0000  Twitter for iPhone
         1891  667832474953625600  2015-11-20 22:30:44 +0000  Twitter for iPhone
         1920  667369227918143488  2015-11-19 15:49:57 +0000  Twitter for iPhone
         1948  666786068205871104  2015-11-18 01:12:41 +0000  Twitter for iPhone
         1945  666826780179869698  2015-11-18 03:54:28 +0000  Twitter for iPhone
         1944  666835007768551424  2015-11-18 04:27:09 +0000  Twitter for iPhone
         1943  666837028449972224  2015-11-18 04:35:11 +0000  Twitter for iPhone
         1939  667044094246576128  2015-11-18 18:17:59 +0000  Twitter for iPhone
         1937  667065535570550784  2015-11-18 19:43:11 +0000  Twitter for iPhone
         1936  667073648344346624  2015-11-18 20:15:26 +0000  Twitter for iPhone
         1933  667138269671505920  2015-11-19 00:32:12 +0000  Twitter for iPhone
         1927  667176164155375616  2015-11-19 03:02:47 +0000  Twitter for iPhone
         1926  667177989038297088  2015-11-19 03:10:02 +0000  Twitter for iPhone
         1924  667188689915760640  2015-11-19 03:52:34 +0000  Twitter for iPhone
         1923  667192066997374976  2015-11-19 04:05:59 +0000  Twitter for iPhone
         1888  667873844930215936  2015-11-21 01:15:07 +0000  Twitter for iPhone
         1883  667911425562669056  2015-11-21 03:44:27 +0000  Twitter for iPhone
         1880  667937095915278337  2015-11-21 05:26:27 +0000  Twitter for iPhone
         1878  668142349051129856  2015-11-21 19:02:04 +0000  Twitter for iPhone
         1877  668154635664932864  2015-11-21 19:50:53 +0000  Twitter for iPhone
         ...                  ...                        ...                 ...
         5219  688916208532455424  2016-01-18 02:49:58 +0000  Twitter for iPhone
         5051  703268521220972544  2016-02-26 17:20:56 +0000  Twitter for iPhone
         5048  703407252292673536  2016-02-27 02:32:12 +0000  Twitter for iPhone
         4771  743222593470234624  2016-06-15 23:24:09 +0000  Twitter for iPhone
         4774  742465774154047488  2016-06-13 21:16:49 +0000  Twitter for iPhone
         5488  675740360753160193  2015-12-12 18:13:51 +0000  Twitter for iPhone
         5227  688519176466644993  2016-01-17 00:32:18 +0000  Twitter for iPhone
         4290  824325613288833024  2017-01-25 18:38:36 +0000  Twitter for iPhone
         4783  741067306818797568  2016-06-10 00:39:48 +0000  Twitter for iPhone
         5477  675898130735476737  2015-12-13 04:40:46 +0000  Twitter for iPhone
```

```
5036  704761120771465216   2016-03-01 20:11:59 +0000   Twitter for iPhone
5234  687818504314159109   2016-01-15 02:08:05 +0000   Twitter for iPhone
5033  704859558691414016   2016-03-02 02:43:09 +0000   Twitter for iPhone
5468  676263575653122048   2015-12-14 04:52:55 +0000   Twitter for iPhone
4037  881536004380872706   2017-07-02 15:32:16 +0000   Twitter for iPhone
5540  674638615994089473   2015-12-09 17:15:54 +0000   Twitter for iPhone
5070  700864154249383937   2016-02-20 02:06:50 +0000   Twitter for iPhone
4249  831315979191906304   2017-02-14 01:35:49 +0000   Twitter Web Client
5505  675334060156301312   2015-12-11 15:19:21 +0000   Twitter for iPhone
5078  700151421916807169   2016-02-18 02:54:41 +0000   Twitter for iPhone
5200  690015576308211712   2016-01-21 03:38:27 +0000   Twitter for iPhone
5543  674447403907457024   2015-12-09 04:36:06 +0000   Twitter for iPhone
6680  752519690950500352   2016-07-11 15:07:30 +0000   Twitter for iPhone
6130  855851453814013952   2017-04-22 18:31:02 +0000   Twitter for iPhone
6291  822872901745569793   2017-01-21 18:26:02 +0000   Twitter for iPhone
5994  889665388333682689   2017-07-25 01:55:32 +0000   Twitter for iPhone
6116  859607811541651456   2017-05-03 03:17:27 +0000   Twitter for iPhone
6661  756275833623502848   2016-07-21 23:53:04 +0000   Twitter for iPhone
6516  780931614150983680   2016-09-28 00:46:20 +0000   Twitter for iPhone
6279  825535076884762624   2017-01-29 02:44:34 +0000   Twitter for iPhone

                                                              text  \
1917              Ermergerd 12/10 https://t.co/PQni2sjPsm
1916  Never seen this breed before. Very pointy pup...
1915  Exotic dog here. Long neck. Weird paws. Obsess...
1911  Two dogs in this one. Both are rare Jujitsu Py...
1906  Another topnotch dog. His name is Big Jumpy Ra...
1901  Here is George. George took a selfie of his ne...
1900  Never seen dog like this. Breathes heavy. Tilt...
1899  What a dog to start the day with. Very calm. L...
1895  Super rare dog. Endangered (?). Thinks it's fu...
1894  Dogs only please. Small cows and other non can...
1893               OMIGOD 12/10 https://t.co/SVMF4Frf1w
1891  THE EYES 12/10\n\nI'm sorry. These are suppose...
1920  Here we have a neat pup. Very white. Cool shad...
1948  Unfamiliar with this breed. Ears pointy af. Wo...
1945  12/10 simply brilliant pup https://t.co/V6ZzG4...
1944  These are Peruvian Feldspars. Their names are ...
1943  My goodness. Very rare dog here. Large. Tail d...
1939            12/10 gimme now https://t.co/QZAnwgnOMB
1937  Here we have a Hufflepuff. Loves vest. Eyes wi...
1936  Here is Dave. He is actually just a skinny leg...
1933  Extremely intelligent dog here. Has learned to...
1927  These are strange dogs. All have toupees. Long...
1926  This is a Dasani Kingfisher from Maine. His na...
1924  Quite an advanced dog here. Impressively dress...
1923  *takes several long deep breaths* omg omg oMG ...
1888  Neat dog. Lots of spikes. Always in push-up po...
```

```
1883  Wow. Armored dog here. Ready for battle. Face ...
1880  This dog resembles a baked potato. Bed looks u...
1878  This lil pup is Oliver. Hops around. Has wings...
1877  Fun dogs here. Top one clearly an athlete. Bot...
...                                                    ...
5219  This pupper just wants to say hello. 11/10 wou...
5051  Happy Friday here's a sleepy pupper 12/10 http...
5048  This pupper doesn't understand gates. 10/10 so...
4771  This is a very rare Great Alaskan Bush Pupper...
4774  Was just informed about this hero pupper and o...
5488  Here's a pupper licking in slow motion. 12/10 ...
5227  This pupper is sprouting a flower out of her h...
4290  Retweet the h*ck out of this 13/10 pupper #Bel...
4783  This is just downright precious af. 12/10 for ...
5477  I'm sure you've all seen this pupper. Not prep...
5036  This pupper killed this great white in an epic...
5234  With great pupper comes great responsibility. ...
5033  Here is a heartbreaking scene of an incredible...
5468  All this pupper wanted to do was go skiing. No...
4037  Here is a pupper approaching maximum borkdrive...
5540  This pupper is fed up with being tickled. 12/1...
5070  "Pupper is a present to world. Here is a bow f...
4249  I couldn't make it to the #WKCDogShow BUT I ha...
5505  Good morning here's a grass pupper. 12/10 http...
5078  If a pupper gave that to me I'd probably start...
5200  This pupper can only sleep on shoes. It's a cr...
5543  This pupper just wants a belly rub. This puppe...
6680  Hopefully this puppo on a swing will help get ...
6130  Here's a puppo participating in the #ScienceMa...
6291  Here's a super supportive puppo participating ...
5994  Here's a puppo that seems to be on the fence a...
6116  Sorry for the lack of posts today. I came home...
6661  When ur older siblings get to play in the deep...
6516  I want to finally rate this iconic puppo who t...
6279  Here's a very loving and accepting puppo. Appe...

                                  expanded_urls     name  \
1917  https://twitter.com/dog_rates/status/667435689...    None
1916  https://twitter.com/dog_rates/status/667437278...    None
1915  https://twitter.com/dog_rates/status/667443425...    None
1911  https://twitter.com/dog_rates/status/667491009...    None
1906  https://twitter.com/dog_rates/status/667524857...    None
1901  https://twitter.com/dog_rates/status/667546741...  George
1900  https://twitter.com/dog_rates/status/667549055...    None
1899  https://twitter.com/dog_rates/status/667724302...    None
1895  https://twitter.com/dog_rates/status/667782464...    None
1894  https://twitter.com/dog_rates/status/667793409...    None
1893  https://twitter.com/dog_rates/status/667801013...    None
```

```
1891    https://twitter.com/dog_rates/status/667832474...        None
1920    https://twitter.com/dog_rates/status/667369227...        None
1948    https://twitter.com/dog_rates/status/666786068...        None
1945    https://twitter.com/dog_rates/status/666826780...        None
1944    https://twitter.com/dog_rates/status/666835007...        None
1943    https://twitter.com/dog_rates/status/666837028...        None
1939    https://twitter.com/dog_rates/status/667044094...        None
1937    https://twitter.com/dog_rates/status/667065535...        None
1936    https://twitter.com/dog_rates/status/667073648...        Dave
1933    https://twitter.com/dog_rates/status/667138269...        None
1927    https://twitter.com/dog_rates/status/667176164...        None
1926    https://twitter.com/dog_rates/status/667177989...           a
1924    https://twitter.com/dog_rates/status/667188689...        None
1923    https://twitter.com/dog_rates/status/667192066...        None
1888    https://twitter.com/dog_rates/status/667873844...        None
1883    https://twitter.com/dog_rates/status/667911425...        None
1880    https://twitter.com/dog_rates/status/667937095...        None
1878    https://twitter.com/dog_rates/status/668142349...        None
1877    https://twitter.com/dog_rates/status/668154635...        None
...                                                      ...       ...
5219    https://twitter.com/dog_rates/status/688916208...        None
5051    https://twitter.com/dog_rates/status/703268521...        None
5048    https://twitter.com/dog_rates/status/703407252...        None
4771    https://twitter.com/dog_rates/status/743222593...           a
4774    https://twitter.com/dog_rates/status/742465774...        None
5488    https://twitter.com/dog_rates/status/675740360...        None
5227    https://twitter.com/dog_rates/status/688519176...        None
4290    https://twitter.com/dog_rates/status/824325613...        None
4783    https://twitter.com/dog_rates/status/741067306...        just
5477    https://twitter.com/dog_rates/status/675898130...        None
5036    https://twitter.com/dog_rates/status/704761120...        None
5234    https://twitter.com/dog_rates/status/687818504...        None
5033    https://twitter.com/dog_rates/status/704859558...           a
5468    https://twitter.com/dog_rates/status/676263575...        None
4037    https://twitter.com/dog_rates/status/881536004...           a
5540    https://twitter.com/dog_rates/status/674638615...        None
5070    https://twitter.com/dog_rates/status/700864154...           a
4249    https://twitter.com/dog_rates/status/831315979...        None
5505    https://twitter.com/dog_rates/status/675334060...        None
5078    https://twitter.com/dog_rates/status/700151421...        None
5200    https://twitter.com/dog_rates/status/690015576...        None
5543    https://twitter.com/dog_rates/status/674447403...        None
6680    https://twitter.com/dog_rates/status/752519690...        None
6130    https://twitter.com/dog_rates/status/855851453...        None
6291    https://twitter.com/dog_rates/status/822872901...        None
5994    https://twitter.com/dog_rates/status/889665388...        None
6116    https://twitter.com/dog_rates/status/859607811...        None
6661    https://twitter.com/dog_rates/status/756275833...        None
```

```
6516   https://twitter.com/dog_rates/status/780931614...      None
6279   https://twitter.com/dog_rates/status/825535076...      None


                                              jpg_url   img_num  \
1917   https://pbs.twimg.com/media/CUM1OHCW4AEgGSi.jpg         1
1916   https://pbs.twimg.com/media/CUM2qWaWoAUZO6L.jpg         1
1915   https://pbs.twimg.com/media/CUM8QZwW4AAVsBl.jpg         1
1911   https://pbs.twimg.com/media/CUNniSlUYAEj1Jl.jpg         1
1906   https://pbs.twimg.com/media/CUOGUfJW4AA_eni.jpg         1
1901   https://pbs.twimg.com/media/CUOaOWXWcAAO_Jy.jpg         1
1900   https://pbs.twimg.com/media/CUOcVCwWsAERUKY.jpg         1
1899   https://pbs.twimg.com/media/CUQ7tv3W4AA3KlI.jpg         1
1895   https://pbs.twimg.com/media/CURwm3cUkAARcO6.jpg         1
1894   https://pbs.twimg.com/media/CUR6jqVWsAEgGot.jpg         1
1893   https://pbs.twimg.com/media/CUSBemVUEAAn-6V.jpg         1
1891   https://pbs.twimg.com/media/CUSeGFNW4AAyyHC.jpg         1
1920   https://pbs.twimg.com/media/CUL4xR9UkAEdlJ6.jpg         1
1948   https://pbs.twimg.com/media/CUDmZIkWcAAIPPe.jpg         1
1945   https://pbs.twimg.com/media/CUELaONUkAAscGC.jpg         1
1944   https://pbs.twimg.com/media/CUES51dXIAEahyG.jpg         1
1943   https://pbs.twimg.com/media/CUEUva1WsAA2jPb.jpg         1
1939   https://pbs.twimg.com/media/CUHREBXXAAE6A9b.jpg         1
1937   https://pbs.twimg.com/media/CUHkkJpXIAA2w3n.jpg         1
1936   https://pbs.twimg.com/media/CUHr8WbWEAEBPgf.jpg         1
1933   https://pbs.twimg.com/media/CUImtzEVAAAZNJo.jpg         1
1927   https://pbs.twimg.com/media/CUJJLtWWsAE-go5.jpg         1
1926   https://pbs.twimg.com/media/CUJK18UWEAEg7AR.jpg         1
1924   https://pbs.twimg.com/media/CUJUk2iWUAAVtOv.jpg         1
1923   https://pbs.twimg.com/media/CUJXpRBXIAANOyz.jpg         1
1888   https://pbs.twimg.com/media/CUTDtyGXIAARxus.jpg         1
1883   https://pbs.twimg.com/media/CUTl5m1WUAAabZG.jpg         1
1880   https://pbs.twimg.com/media/CUT9PuQWwAABQv7.jpg         1
1878   https://pbs.twimg.com/media/CUW37BzWsAAlJlN.jpg         1
1877   https://pbs.twimg.com/media/CUXDGR2WcAAUQKz.jpg         1
...                                               ...       ...
5219   https://pbs.twimg.com/media/CY-Fn1FWEAQhzhs.jpg         1
5051   https://pbs.twimg.com/media/CcKC-5LW4AAK-nb.jpg         1
5048   https://pbs.twimg.com/media/CcMBJODUsAI5-A9.jpg         1
4771   https://pbs.twimg.com/media/ClBO9zOWYAAA1jz.jpg         1
4774   https://pbs.twimg.com/media/Ck3EribXEAAPhZn.jpg         1
5488   https://pbs.twimg.com/ext_tw_video_thumb/67574...         1
5227   https://pbs.twimg.com/media/CY4ciRFUMAAovos.jpg         1
4290   https://pbs.twimg.com/media/C3CXxaoWQAAiLuC.jpg         1
4783   https://pbs.twimg.com/media/CkjMx99UoAM2B1a.jpg         1
5477   https://pbs.twimg.com/media/CWFFt3_XIAArIYK.jpg         1
5036   https://pbs.twimg.com/media/CcfQgHVWoAAxauy.jpg         1
5234   https://pbs.twimg.com/media/CYufR8_WQAAWCqo.jpg         1
5033   https://pbs.twimg.com/media/CcgqBNVW8AE76lv.jpg         1
```

```
5468        https://pbs.twimg.com/media/CWKSIfUUYAAiOBO.jpg          1
4037   https://pbs.twimg.com/ext_tw_video_thumb/88153...          1
5540        https://pbs.twimg.com/media/CVzMPh1UsAELQ_p.jpg          1
5070        https://pbs.twimg.com/media/Cbn4OqKWwAADGWt.jpg          1
4249        https://pbs.twimg.com/media/C4lst0bXAAE6MP8.jpg          4
5505        https://pbs.twimg.com/media/CV9EvZNUwAAgLCK.jpg          2
5078        https://pbs.twimg.com/media/CbdwATgWwAABGID.jpg          1
5200        https://pbs.twimg.com/media/CZNtgWhWkAAbq3W.jpg          2
5543        https://pbs.twimg.com/media/CVweVUfW4AACPwI.jpg          1
6680        https://pbs.twimg.com/media/CnF8qVDWYAAhOg1.jpg          3
6130        https://pbs.twimg.com/media/C-CYWrvWAAU8AXH.jpg          1
6291        https://pbs.twimg.com/media/C2tugXLXgAArJO4.jpg          1
5994        https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg          1
6116        https://pbs.twimg.com/media/C-3wvtxXcAUTuBE.jpg          1
6661        https://pbs.twimg.com/media/Cn7U2xlW8AI9Pqp.jpg          1
6516        https://pbs.twimg.com/media/CtZtJxAXEAAyPGd.jpg          1
6279        https://pbs.twimg.com/media/C3TjvitXAAAI-QH.jpg          1


                                  p1    p1_conf    ...      p3_conf  p3_dog  \
1917                     Rottweiler  0.999091    ...     0.000157    True
1916                      porcupine  0.989154    ...     0.000966   False
1915                          goose  0.980815    ...     0.005255   False
1911                         borzoi  0.852088    ...     0.005730   False
1906                           hare  0.447893    ...     0.088122    True
1901                     toy_poodle  0.787424    ...     0.004047   False
1900                   electric_fan  0.984377    ...     0.001901   False
1899                           ibex  0.619098    ...     0.074673   False
1895                        lorikeet  0.466149    ...     0.054247   False
1894                       dalmatian  0.535073    ...     0.008164    True
1893           flat-coated_retriever  0.508392    ...     0.048920    True
1891             miniature_pinscher  0.214200    ...     0.104152    True
1920                          teddy  0.709545    ...     0.028568   False
1948                          snail  0.999888    ...     0.000026   False
1945                    Maltese_dog  0.359383    ...     0.106007    True
1944                        Airedale  0.448459    ...     0.110183   False
1943                     triceratops  0.442113    ...     0.043255   False
1939               golden_retriever  0.765266    ...     0.010667   False
1937                   jigsaw_puzzle  0.560001    ...     0.042568   False
1936                       Chihuahua  0.483682    ...     0.057495    True
1933      West_Highland_white_terrier  0.747713    ...     0.001804    True
1927      soft-coated_wheaten_terrier  0.318981    ...     0.106014    True
1926                          vizsla  0.259249    ...     0.112369    True
1924                          vacuum  0.335830    ...     0.140703   False
1923                      Rottweiler  0.283640    ...     0.095585    True
1888                   common_iguana  0.999647    ...     0.000128   False
1883                   frilled_lizard  0.257695    ...     0.085317   False
1880                         hamster  0.172078    ...     0.059995   False
1878                          Angora  0.918834    ...     0.011015   False
```

1877	Arctic_fox	0.473584	...	0.080948	False
...	...	...	...	...	...
5219	Pembroke	0.430544	...	0.154352	True
5051	wool	0.525434	...	0.038243	True
5048	doormat	0.201058	...	0.082924	False
4771	kuvasz	0.350629	...	0.087662	True
4774	web_site	0.997154	...	0.000268	False
5488	golden_retriever	0.800495	...	0.068415	True
5227	Pembroke	0.696372	...	0.050592	True
4290	Pembroke	0.990793	...	0.000262	True
4783	golden_retriever	0.843799	...	0.035711	True
5477	Labrador_retriever	0.407430	...	0.074597	True
5036	Siamese_cat	0.202294	...	0.072097	True
5234	Lakeland_terrier	0.873029	...	0.017031	True
5033	pug	0.284428	...	0.138915	False
5468	teddy	0.098283	...	0.077852	False
4037	Samoyed	0.281463	...	0.114854	False
5540	Pomeranian	0.846986	...	0.002605	True
5070	kuvasz	0.805857	...	0.003491	True
4249	briard	0.982755	...	0.004693	True
5505	Pembroke	0.773135	...	0.039036	True
5078	tennis_ball	0.328236	...	0.134080	True
5200	malamute	0.949609	...	0.016663	True
5543	Brabancon_griffon	0.409909	...	0.074820	True
6680	swing	0.999984	...	0.000001	True
6130	flat-coated_retriever	0.321676	...	0.096100	True
6291	Lakeland_terrier	0.196015	...	0.069126	True
5994	Pembroke	0.966327	...	0.004633	True
6116	golden_retriever	0.895529	...	0.019285	True
6661	Airedale	0.602957	...	0.086276	True
6516	padlock	0.731564	...	0.036469	False
6279	Rottweiler	0.681495	...	0.024525	True

	favorites	retweets	user_followers	dog_stage	rating_numerator \
1917	310	85	6992426	None	12
1916	458	248	6992426	None	10
1915	803	604	6992426	None	6
1911	547	235	6992426	None	7.5
1906	1745	1162	6992426	None	12
1901	341	130	6992426	None	9
1900	5961	2386	6993394	None	1
1899	502	332	6992426	None	7
1895	423	253	6992426	None	9
1894	718	341	6992426	None	8
1893	338	98	6992426	None	12
1891	294	65	6992426	None	12
1920	375	168	6992426	None	10
1948	775	496	6992425	None	2

1945	258	98	6992425	None	12
1944	218	80	6992426	None	10
1943	833	566	6992426	None	3
1939	188	51	6992426	None	12
1937	169	49	6992426	None	8
1936	412	126	6992426	None	10
1933	4717	2295	6992426	None	10
1927	624	473	6992426	None	4
1926	193	57	6992426	None	8
1924	759	430	6992426	None	10
1923	396	109	6992426	None	12
1888	649	426	6992426	None	10
1883	510	316	6992427	None	5
1880	1313	831	6992427	None	3
1878	579	300	6992427	None	2
1877	511	330	6992427	None	9
...	...	...	...	...	...
5219	2918	954	6992421	pupper	11
5051	2104	607	6992419	pupper	12
5048	2632	761	6993387	pupper	10
4771	6620	2090	6992417	pupper	12
4774	7750	4243	6992417	pupper	14
5488	1217	373	6992427	pupper	12
5227	2466	798	6992422	pupper	12
4290	12692	11557	6992407	pupper	13
4783	10074	3385	6992417	pupper	12
5477	1718	623	6992427	pupper	10
5036	7085	3167	6992419	pupper	13
5234	2678	1043	6992422	pupper	12
5033	2405	597	6992419	pupper	10
5468	2183	590	6992427	pupper	10
4037	49361	16076	6992403	pupper	14
5540	1741	621	6992426	pupper	12
5070	2768	667	6992419	pupper	12
4249	6974	1230	6992406	pupper	13
5505	2915	1387	6992426	pupper	12
5078	2393	723	6992418	pupper	11
5200	2681	807	6992420	pupper	12
5543	1103	381	6992426	pupper	10
6680	7939	3790	6992416	puppo	11
6130	46856	18650	6992403	puppo	13
6291	142980	48957	6992407	puppo	13
5994	48011	10108	6992403	puppo	13
6116	19036	1640	6992403	puppo	13
6661	6941	1697	6992416	puppo	10
6516	23535	8237	6993378	puppo	13
6279	56057	19124	6992407	puppo	14

```
      rating_denominator  dogs_count  dog_name
1917                  10           1       NaN
1916                  10           1       NaN
1915                  10           1       NaN
1911                  10           2       NaN
1906                  10           1       NaN
1901                  10           1       NaN
1900                  10           1       NaN
1899                  10           1       NaN
1895                  10           1       NaN
1894                  10           1       NaN
1893                  10           1       NaN
1891                  10           1       NaN
1920                  10           1       NaN
1948                  10           1       NaN
1945                  10           1       NaN
1944                  10           2       NaN
1943                  10           1       NaN
1939                  10           1       NaN
1937                  10           1       NaN
1936                  10           1       NaN
1933                  10           1       NaN
1927                  10           1       NaN
1926                  10           1       NaN
1924                  10           1       NaN
1923                  10           1       NaN
1888                  10           1       NaN
1883                  10           1       NaN
1880                  10           1       NaN
1878                  10           1       NaN
1877                  10           1       NaN
...                  ...         ...       ...
5219                  10           1       NaN
5051                  10           1       NaN
5048                  10           1       NaN
4771                  10           1       NaN
4774                  10           1       NaN
5488                  10           1       NaN
5227                  10           1       NaN
4290                  10           1       NaN
4783                  10           1       NaN
5477                  10           1       NaN
5036                  10           1       NaN
5234                  10           1       NaN
5033                  10           1       NaN
5468                  10           1       NaN
4037                  10           1       NaN
5540                  10           1       NaN
```

```
5070                       10              1           NaN
4249                       10              1           NaN
5505                       10              1           NaN
5078                       10              1           NaN
5200                       10              1           NaN
5543                       10              1           NaN
6680                       10              1           NaN
6130                       10              1           NaN
6291                       10              1           NaN
5994                       10              1           NaN
6116                       10              1           NaN
6661                       10              1           NaN
6516                       10              1           NaN
6279                       10              1           NaN

[625 rows x 25 columns]

In [55]: df_master = df_master.drop(['name'], 1)


         list(df_master)

Out[55]: ['tweet_id',
          'timestamp',
          'source',
          'text',
          'expanded_urls',
          'jpg_url',
          'img_num',
          'p1',
          'p1_conf',
          'p1_dog',
          'p2',
          'p2_conf',
          'p2_dog',
          'p3',
          'p3_conf',
          'p3_dog',
          'favorites',
          'retweets',
          'user_followers',
          'dog_stage',
          'rating_numerator',
          'rating_denominator',
          'dogs_count',
          'dog_name']

In [56]: male = ['He', 'he', 'him', 'his', "he's", 'himself']
         female = ['She', 'she', 'her', 'hers', 'herself', "she's"]
```

```python
        dog_gender = []

        for text in df_master['text']:
            # Male
            if any(map(lambda v:v in male, text.split())):
                dog_gender.append('male')
            # Female
            elif any(map(lambda v:v in female, text.split())):
                dog_gender.append('female')
            # If group or not specified
            else:
                dog_gender.append('NaN')

        len(dog_gender)


        df_master['dog_gender'] = dog_gender

        # Test
        print("dog_gender count \n", df_master.dog_gender.value_counts())
```

```
dog_gender count
 NaN        1132
male        636
female      226
Name: dog_gender, dtype: int64
```

```python
In [58]: df_master.loc[df_master['dog_name'] == 'NaN', 'dog_name'] = None
         df_master.loc[df_master['dog_gender'] == 'NaN', 'dog_gender'] = None
         df_master.loc[df_master['rating_numerator'] == 'NaN', 'rating_numerator'] = 0
         df_master.loc[df_master['rating_denominator'] == 'NaN', 'rating_denominator'] = 0

In [59]: df_master.dtypes

Out[59]: tweet_id                int64
         timestamp              object
         source                 object
         text                   object
         expanded_urls          object
         jpg_url                object
         img_num                 int64
         p1                     object
         p1_conf               float64
         p1_dog                   bool
         p2                     object
         p2_conf               float64
         p2_dog                   bool
```

```
          p3                            object
          p3_conf                      float64
          p3_dog                          bool
          favorites                     int64
          retweets                      int64
          user_followers                int64
          dog_stage                    object
          rating_numerator             object
          rating_denominator           object
          dogs_count                   object
          dog_name                     object
          dog_gender                   object
          dtype: object
```

In [60]: df_master['tweet_id'] = df_master['tweet_id'].astype(object)
         df_master['timestamp'] = pd.to_datetime(df_master.timestamp)
         df_master['source'] = df_master['source'].astype('category')
         df_master['favorites'] = df_master['favorites'].astype(int)
         df_master['retweets'] = df_master['retweets'].astype(int)
         df_master['user_followers'] = df_master['user_followers'].astype(int)
         df_master['dog_stage'] = df_master['dog_stage'].astype('category')
         df_master['rating_numerator'] = df_master['rating_numerator'].astype(float)
         df_master['rating_denominator'] = df_master['rating_denominator'].astype(float)
         df_master['dogs_count'] = df_master['dogs_count'].astype(int)
         df_master['dog_gender'] = df_master['dog_gender'].astype('category')

In [61]: df_master.dtypes

Out[61]: tweet_id   object
         timestamp   datetime64[ns]
         source   category
         text   object
         expanded_urls   object
         jpg_url   object
         img_num   int64
         p1   object
         p1_conf   float64
         p1_dog   bool
         p2   object
         p2_conf   float64
         p2_dog   bool
         p3   object
         p3_conf   float64
         p3_dog   bool
         favorites   int64
         retweets   int64
         user_followers   int64
         dog_stage   category
```

```
         rating_numerator             float64
         rating_denominator           float64
         dogs_count                     int64
         dog_name                      object
         dog_gender                  category
         dtype: object

In [62]: print(df_master.rating_denominator.value_counts())
         df_master.drop('rating_denominator', 1, inplace = True)

10.0     1981
80.0        2
50.0        2
170.0       1
150.0       1
130.0       1
70.0        1
0.0         1
120.0       1
40.0        1
90.0        1
110.0       1
Name: rating_denominator, dtype: int64


In [63]: list(df_master)

Out[63]: ['tweet_id',
          'timestamp',
          'source',
          'text',
          'expanded_urls',
          'jpg_url',
          'img_num',
          'p1',
          'p1_conf',
          'p1_dog',
          'p2',
          'p2_conf',
          'p2_dog',
          'p3',
          'p3_conf',
          'p3_dog',
          'favorites',
          'retweets',
          'user_followers',
          'dog_stage',
          'rating_numerator',
          'dogs_count',
```

```
            'dog_name',
            'dog_gender']

In [64]: df_master.to_csv('twitter_archive_master.csv', index=False, encoding = 'utf-8')

In [65]: df_master = pd.read_csv('twitter_archive_master.csv')
         df_master.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1994 entries, 0 to 1993
Data columns (total 24 columns):
tweet_id           1994 non-null int64
timestamp          1994 non-null object
source             1994 non-null object
text               1994 non-null object
expanded_urls      1994 non-null object
jpg_url            1994 non-null object
img_num            1994 non-null int64
p1                 1994 non-null object
p1_conf            1994 non-null float64
p1_dog             1994 non-null bool
p2                 1994 non-null object
p2_conf            1994 non-null float64
p2_dog             1994 non-null bool
p3                 1994 non-null object
p3_conf            1994 non-null float64
p3_dog             1994 non-null bool
favorites          1994 non-null int64
retweets           1994 non-null int64
user_followers     1994 non-null int64
dog_stage          1994 non-null object
rating_numerator   1994 non-null float64
dogs_count         1994 non-null int64
dog_name           1369 non-null object
dog_gender          862 non-null object
dtypes: bool(3), float64(4), int64(6), object(11)
memory usage: 333.1+ KB


In [66]: df_master

Out[66]:                 tweet_id             timestamp             source  \
         0      667405339315146752   2015-11-19 18:13:27   Twitter for iPhone
         1      667435689202614272   2015-11-19 20:14:03   Twitter for iPhone
         2      667437278097252352   2015-11-19 20:20:22   Twitter for iPhone
         3      667443425659232256   2015-11-19 20:44:47   Twitter for iPhone
         4      667453023279554560   2015-11-19 21:22:56   Twitter Web Client
         5      667455448082227200   2015-11-19 21:32:34   Twitter Web Client
         6      667470559035432960   2015-11-19 22:32:36   Twitter Web Client
```

```
7       6674910093796065528    2015-11-19 23:53:52    Twitter Web Client
8       667495797102141441     2015-11-20 00:12:54    Twitter Web Client
9       667502640335572993     2015-11-20 00:40:05    Twitter Web Client
10      667509364010450944     2015-11-20 01:06:48    Twitter Web Client
11      667517642048163840     2015-11-20 01:39:42    Twitter Web Client
12      667524857454854144     2015-11-20 02:08:22    Twitter Web Client
13      667530908589760512     2015-11-20 02:32:25    Twitter Web Client
14      667534815156183040     2015-11-20 02:47:56    Twitter Web Client
15      667538891197542400     2015-11-20 03:04:08    Twitter Web Client
16      667544320556335104     2015-11-20 03:25:43    Twitter Web Client
17      667546741521195010     2015-11-20 03:35:20    Twitter Web Client
18      667549055577362432     2015-11-20 03:44:31    Twitter Web Client
19      667724302356258817     2015-11-20 15:20:54    Twitter Web Client
20      667728196545200128     2015-11-20 15:36:22    Twitter Web Client
21      667766675769573376     2015-11-20 18:09:16    Twitter Web Client
22      667773195014021121     2015-11-20 18:35:10    Twitter Web Client
23      667782464991965184     2015-11-20 19:12:01    Twitter for iPhone
24      667793409583771648     2015-11-20 19:55:30    Twitter for iPhone
25      667801013445750784     2015-11-20 20:25:43    Twitter for iPhone
26      667806454573760512     2015-11-20 20:47:20    Twitter for iPhone
27      667393430834667520     2015-11-19 17:26:08    Twitter for iPhone
28      667832474953625600     2015-11-20 22:30:44    Twitter for iPhone
29      667369227918143488     2015-11-19 15:49:57    Twitter for iPhone
...            ...                     ...                    ...
1964    690015576308211712     2016-01-21 03:38:27    Twitter for iPhone
1965    675113801096802304     2015-12-11 00:44:07    Twitter for iPhone
1966    689905486972461056     2016-01-20 20:21:00    Twitter for iPhone
1967    874296783580663808     2017-06-12 16:06:11    Twitter for iPhone
1968    674447403907457024     2015-12-09 04:36:06    Twitter for iPhone
1969    702598099714314240     2016-02-24 20:56:55    Twitter for iPhone
1970    689623661272240129     2016-01-20 01:41:08    Twitter for iPhone
1971    825026590719483904     2017-01-27 17:04:02    Twitter for iPhone
1972    889531135344209921     2017-07-24 17:02:04    Twitter for iPhone
1973    752519690950500352     2016-07-11 15:07:30    Twitter for iPhone
1974    855851453814013952     2017-04-22 18:31:02    Twitter for iPhone
1975    822872901745569793     2017-01-21 18:26:02    Twitter for iPhone
1976    751132876104687617     2016-07-07 19:16:47    Twitter for iPhone
1977    793195938047070209     2016-10-31 21:00:23    Twitter for iPhone
1978    874012996292530176     2017-06-11 21:18:31    Twitter for iPhone
1979    889665388333682689     2017-07-25 01:55:32    Twitter for iPhone
1980    859607811541651456     2017-05-03 03:17:27    Twitter for iPhone
1981    787717603741622272     2016-10-16 18:11:26    Twitter for iPhone
1982    819952236453363712     2017-01-13 17:00:21    Twitter for iPhone
1983    738537504001953792     2016-06-03 01:07:16    Twitter for iPhone
1984    744995568523612160     2016-06-20 20:49:19    Twitter for iPhone
1985    756275833623502848     2016-07-21 23:53:04    Twitter for iPhone
1986    780931614150983680     2016-09-28 00:46:20    Twitter for iPhone
1987    802239329049477120     2016-11-25 19:55:35    Twitter for iPhone
```

```
1988  803773340896923648  2016-11-30 01:31:12  Twitter for iPhone
1989  790946055508652032  2016-10-25 16:00:09  Twitter for iPhone
1990  867421006826221569  2017-05-24 16:44:18  Twitter for iPhone
1991  878776093423087618  2017-06-25 00:45:22  Twitter for iPhone
1992  825535076884762624  2017-01-29 02:44:34  Twitter for iPhone
1993  743253157753532416  2016-06-16 01:25:36  Twitter for iPhone

                                                             text  \
0     This is Biden. Biden just tripped... 7/10 http...
1                    Ermergerd 12/10 https://t.co/PQni2sjPsm
2     Never seen this breed before. Very pointy pup...
3     Exotic dog here. Long neck. Weird paws. Obsess...
4     Meet Cupcake. I would do unspeakable things fo...
5     This is Reese and Twips. Reese protects Twips...
6     This is a northern Wahoo named Kohl. He runs t...
7     Two dogs in this one. Both are rare Jujitsu Py...
8     This is Philippe from Soviet Russia. Commandin...
9     Say hello to Hall and Oates. Oates is winking ...
10    This a Norwegian Pewterschmidt named Tickles. ...
11    This is Dook &amp; Milo. Dook is struggling to...
12    Another topnotch dog. His name is Big Jumpy Ra...
13    Meet Naphaniel. He doesn't necessarily enjoy h...
14    This is Frank (pronounced "Fronq"). Too many b...
15    This is a southwest Coriander named Klint. Hat...
16    This is Kial. Kial is either wearing a cape, w...
17    Here is George. George took a selfie of his ne...
18    Never seen dog like this. Breathes heavy. Tilt...
19    What a dog to start the day with. Very calm. L...
20    Meet Olive. He comes to spot by tree to remini...
21    This is Calvin. He is a Luxembourgian Mayo. Ha...
22    This is a rare Hungarian Pinot named Jessiga. ...
23    Super rare dog. Endangered (?). Thinks it's fu...
24    Dogs only please. Small cows and other non can...
25                   OMIGOD 12/10 https://t.co/SVMF4Frf1w
26    This is Filup. He is overcome with joy after f...
27    This is Fwed. He is a Canadian Asian Taylormad...
28    THE EYES 12/10\n\nI'm sorry. These are suppose...
29    Here we have a neat pup. Very white. Cool shad...
...                                                    ...
1964  This pupper can only sleep on shoes. It's a cr...
1965  Meet Zuzu. He just graduated college. Astute p...
1966  Say hello to Gizmo. He's quite the pupper. Con...
1967  This is Jed. He may be the fanciest pupper in ...
1968  This pupper just wants a belly rub. This puppe...
1969  This is Sansa. She's gotten too big for her ch...
1970  This is Lucy. She's terrified of the stuffed b...
1971  Say hello to Pablo. He's one gorgeous puppo. A...
1972  This is Stuart. He's sporting his favorite fan...
```

```
1973  Hopefully this puppo on a swing will help get ...
1974  Here's a puppo participating in the #ScienceMa...
1975  Here's a super supportive puppo participating ...
1976  This is Cooper. He's just so damn happy. 10/10...
1977  Say hello to Lily. She's pupset that her costu...
1978  This is Sebastian. He can't see all the colors...
1979  Here's a puppo that seems to be on the fence a...
1980  Sorry for the lack of posts today. I came home...
1981  This is Tonks. She is a service puppo. Can hea...
1982  This is Oliver. He has dreams of being a servi...
1983  This is Bayley. She fell asleep trying to esca...
1984  This is Abby. She got her face stuck in a glas...
1985  When ur older siblings get to play in the deep...
1986  I want to finally rate this iconic puppo who t...
1987  This is Loki. He'll do your taxes for you. Can...
1988  This is Diogi. He fell in the pool as soon as ...
1989  This is Betty. She's assisting with the dishes...
1990  This is Shikha. She just watched you drop a sk...
1991  This is Snoopy. He's a proud #PrideMonthPuppo...
1992  Here's a very loving and accepting puppo. Appe...
1993  This is Kilo. He cannot reach the snackum. Nif...

                                           expanded_urls  \
0      https://twitter.com/dog_rates/status/667405339...
1      https://twitter.com/dog_rates/status/667435689...
2      https://twitter.com/dog_rates/status/667437278...
3      https://twitter.com/dog_rates/status/667443425...
4      https://twitter.com/dog_rates/status/667453023...
5      https://twitter.com/dog_rates/status/667455448...
6      https://twitter.com/dog_rates/status/667470559...
7      https://twitter.com/dog_rates/status/667491009...
8      https://twitter.com/dog_rates/status/667495797...
9      https://twitter.com/dog_rates/status/667502640...
10     https://twitter.com/dog_rates/status/667509364...
11     https://twitter.com/dog_rates/status/667517642...
12     https://twitter.com/dog_rates/status/667524857...
13     https://twitter.com/dog_rates/status/667530908...
14     https://twitter.com/dog_rates/status/667534815...
15     https://twitter.com/dog_rates/status/667538891...
16     https://twitter.com/dog_rates/status/667544320...
17     https://twitter.com/dog_rates/status/667546741...
18     https://twitter.com/dog_rates/status/667549055...
19     https://twitter.com/dog_rates/status/667724302...
20     https://twitter.com/dog_rates/status/667728196...
21     https://twitter.com/dog_rates/status/667766675...
22     https://twitter.com/dog_rates/status/667773195...
23     https://twitter.com/dog_rates/status/667782464...
24     https://twitter.com/dog_rates/status/667793409...
```

```
25      https://twitter.com/dog_rates/status/667801013...
26      https://twitter.com/dog_rates/status/667806454...
27      https://twitter.com/dog_rates/status/667393430...
28      https://twitter.com/dog_rates/status/667832474...
29      https://twitter.com/dog_rates/status/667369227...
...                                                  ...
1964    https://twitter.com/dog_rates/status/690015576...
1965    https://twitter.com/dog_rates/status/675113801...
1966    https://twitter.com/dog_rates/status/689905486...
1967    https://twitter.com/dog_rates/status/874296783...
1968    https://twitter.com/dog_rates/status/674447403...
1969    https://twitter.com/dog_rates/status/702598099...
1970    https://twitter.com/dog_rates/status/689623661...
1971    https://www.gofundme.com/my-puppys-double-cata...
1972    https://twitter.com/dog_rates/status/889531135...
1973    https://twitter.com/dog_rates/status/752519690...
1974    https://twitter.com/dog_rates/status/855851453...
1975    https://twitter.com/dog_rates/status/822872901...
1976    https://twitter.com/dog_rates/status/751132876...
1977    https://twitter.com/dog_rates/status/793195938...
1978    https://twitter.com/dog_rates/status/874012996...
1979    https://twitter.com/dog_rates/status/889665388...
1980    https://twitter.com/dog_rates/status/859607811...
1981    https://twitter.com/dog_rates/status/787717603...
1982    https://www.gofundme.com/servicedogoliver,http...
1983    https://twitter.com/dog_rates/status/738537504...
1984    https://twitter.com/dog_rates/status/744995568...
1985    https://twitter.com/dog_rates/status/756275833...
1986    https://twitter.com/dog_rates/status/780931614...
1987    https://twitter.com/dog_rates/status/802239329...
1988    https://twitter.com/dog_rates/status/803773340...
1989    https://twitter.com/dog_rates/status/790946055...
1990    https://twitter.com/dog_rates/status/867421006...
1991    https://twitter.com/dog_rates/status/878776093...
1992    https://twitter.com/dog_rates/status/825535076...
1993    https://twitter.com/dog_rates/status/743253157...

                                          jpg_url  img_num  \
0       https://pbs.twimg.com/media/CUMZnmhUEAEbtis.jpg        1
1       https://pbs.twimg.com/media/CUM1OHCW4AEgGSi.jpg        1
2       https://pbs.twimg.com/media/CUM2qWaWoAUZO6L.jpg        1
3       https://pbs.twimg.com/media/CUM8QZwW4AAVsBl.jpg        1
4       https://pbs.twimg.com/media/CUNE_OSUwAAdHhX.jpg        1
5       https://pbs.twimg.com/media/CUNHMXTU8AAS3HH.jpg        1
6       https://pbs.twimg.com/media/CUNU78YWEAECmpB.jpg        1
7       https://pbs.twimg.com/media/CUNniS1UYAEj1Jl.jpg        1
8       https://pbs.twimg.com/media/CUNr4-7UwAAg2lq.jpg        1
9       https://pbs.twimg.com/media/CUNyHTMUYAAQVch.jpg        1
```

```
10      https://pbs.twimg.com/media/CUN4Or5UAAAa5K4.jpg       1
11      https://pbs.twimg.com/media/CUN_wiBUkAAakTO.jpg       1
12      https://pbs.twimg.com/media/CUOGUfJW4AA_eni.jpg       1
13      https://pbs.twimg.com/media/CUOL0uGUkAAx7yh.jpg       1
14      https://pbs.twimg.com/media/CUOPYI5UcAAj_nO.jpg       1
15      https://pbs.twimg.com/media/CUOTFZOW4AABsfW.jpg       1
16      https://pbs.twimg.com/media/CUOYBbbWIAAXQGU.jpg       1
17      https://pbs.twimg.com/media/CUOaOWXWcAAO_Jy.jpg       1
18      https://pbs.twimg.com/media/CUOcVCwWsAERUKY.jpg       1
19      https://pbs.twimg.com/media/CUQ7tv3W4AA3KlI.jpg       1
20      https://pbs.twimg.com/media/CUQ_QahUAAAVQjn.jpg       1
21      https://pbs.twimg.com/media/CURiQMnUAAAPT2M.jpg       1
22      https://pbs.twimg.com/media/CURoLrOVEAAaWdR.jpg       1
23      https://pbs.twimg.com/media/CURwm3cUkAARcO6.jpg       1
24      https://pbs.twimg.com/media/CUR6jqVWsAEgGot.jpg       1
25      https://pbs.twimg.com/media/CUSBemVUEAAn-6V.jpg       1
26      https://pbs.twimg.com/media/CUSGbXeVAAAgztZ.jpg       1
27      https://pbs.twimg.com/media/CUMOyd3XIAAl13H.jpg       1
28      https://pbs.twimg.com/media/CUSeGFNW4AAyyHC.jpg       1
29      https://pbs.twimg.com/media/CUL4xR9UkAEdlJ6.jpg       1
...                           ...                            ...
1964    https://pbs.twimg.com/media/CZNtgWhWkAAbq3W.jpg       2
1965    https://pbs.twimg.com/media/CV58a4nXAAApywo.jpg       1
1966    https://pbs.twimg.com/media/CZMJYCRVAAE35Wk.jpg       4
1967    https://pbs.twimg.com/media/DCIgSROXgAANEOY.jpg       1
1968    https://pbs.twimg.com/media/CVweVUfW4AACPwI.jpg       1
1969    https://pbs.twimg.com/media/CcAhPevW8AAoknv.jpg       1
1970    https://pbs.twimg.com/media/CZIJD2SWIAMJgNI.jpg       1
1971    https://pbs.twimg.com/media/C3MVTeHWcAAGNfx.jpg       2
1972    https://pbs.twimg.com/media/DFg_2PVW0AEHN3p.jpg       1
1973    https://pbs.twimg.com/media/CnF8qVDWYAAhOg1.jpg       3
1974    https://pbs.twimg.com/media/C-CYWrvWAAU8AXH.jpg       1
1975    https://pbs.twimg.com/media/C2tugXLXgAArJO4.jpg       1
1976    https://pbs.twimg.com/media/CmyPXNOW8AEtaJ-.jpg       1
1977    https://pbs.twimg.com/media/CwH_foYWgAEvTyI.jpg       2
1978    https://pbs.twimg.com/media/DCEeLxjXsAAvNSM.jpg       2
1979    https://pbs.twimg.com/media/DFi579UWsAAatzw.jpg       1
1980    https://pbs.twimg.com/media/C-3wvtxXcAUTuBE.jpg       1
1981    https://pbs.twimg.com/media/Cu6I9vvWIAAZGOa.jpg       3
1982    https://pbs.twimg.com/media/C2EONHNWQAUWxkP.jpg       1
1983    https://pbs.twimg.com/media/Cj_P7rSUgAAYQbz.jpg       1
1984    https://pbs.twimg.com/media/ClbBg4WWEAMjwJu.jpg       1
1985    https://pbs.twimg.com/media/Cn7U2xlW8AI9Pqp.jpg       1
1986    https://pbs.twimg.com/media/CtZtJxAXEAAyPGd.jpg       1
1987    https://pbs.twimg.com/media/CyIgaTEVEAA-9zS.jpg       2
1988    https://pbs.twimg.com/media/CyeTku-XcAALkBd.jpg       2
1989    https://pbs.twimg.com/media/CvoBPWRWgAA4het.jpg       1
1990    https://pbs.twimg.com/media/DAmyy8FXYAIH8Ty.jpg       1
```

```
1991   https://pbs.twimg.com/media/DDIKMXzWOAEibje.jpg           2
1992   https://pbs.twimg.com/media/C3TjvitXAAAI-QH.jpg           1
1993   https://pbs.twimg.com/media/ClCQzFUUYAA5vAu.jpg           1


                                  p1    p1_conf  p1_dog    ...      p3_conf  \
0                    Saint_Bernard  0.381377    True    ...      0.069357
1                       Rottweiler  0.999091    True    ...      0.000157
2                        porcupine  0.989154   False    ...      0.000966
3                            goose  0.980815   False    ...      0.005255
4                Labrador_retriever  0.825670    True    ...      0.054018
5                  Tibetan_terrier  0.676376    True    ...      0.040576
6                       toy_poodle  0.304175    True    ...      0.073316
7                            borzoi  0.852088    True    ...      0.005730
8                        Chihuahua  0.143957    True    ...      0.092482
9                Labrador_retriever  0.996709    True    ...      0.000712
10                           beagle  0.636169    True    ...      0.082549
11               Italian_greyhound  0.125176    True    ...      0.081347
12                            hare  0.447893   False    ...      0.088122
13               golden_retriever  0.633037    True    ...      0.046184
14                        Pembroke  0.435254    True    ...      0.033158
15               Yorkshire_terrier  0.618957    True    ...      0.053412
16                       Pomeranian  0.412893    True    ...      0.071960
17                       toy_poodle  0.787424    True    ...      0.004047
18                     electric_fan  0.984377   False    ...      0.001901
19                             ibex  0.619098   False    ...      0.074673
20                           kuvasz  0.360159    True    ...      0.270673
21                      fire_engine  0.883493   False    ...      0.012773
22   West_Highland_white_terrier  0.360465    True    ...      0.069038
23                         lorikeet  0.466149   False    ...      0.054247
24                        dalmatian  0.535073    True    ...      0.008164
25            flat-coated_retriever  0.508392    True    ...      0.048920
26                          toyshop  0.253089   False    ...      0.112799
27                         papillon  0.557009    True    ...      0.073473
28              miniature_pinscher  0.214200    True    ...      0.104152
29                            teddy  0.709545   False    ...      0.028568
...                            ...       ...     ...    ...          ...
1964                       malamute  0.949609    True    ...      0.016663
1965                            bow  0.168020   False    ...      0.043312
1966                      Pomeranian  0.943331    True    ...      0.007165
1967                   cocker_spaniel  0.437216    True    ...      0.157402
1968               Brabancon_griffon  0.409909    True    ...      0.074820
1969                           kelpie  0.219179    True    ...      0.074440
1970                       toy_poodle  0.279604    True    ...      0.077481
1971                       Eskimo_dog  0.524454    True    ...      0.004976
1972                golden_retriever  0.953442    True    ...      0.007958
1973                            swing  0.999984   False    ...      0.000001
1974           flat-coated_retriever  0.321676    True    ...      0.096100
1975                 Lakeland_terrier  0.196015    True    ...      0.069126
```

75

```
1976              Labrador_retriever  0.929390   True   ...   0.007610
1977              Labrador_retriever  0.654762   True   ...   0.042339
1978                        Cardigan  0.806674   True   ...   0.049182
1979                        Pembroke  0.966327   True   ...   0.004633
1980                golden_retriever  0.895529   True   ...   0.019285
1981                 German_shepherd  0.992339   True   ...   0.000853
1982  American_Staffordshire_terrier  0.925505   True   ...   0.020412
1983                             chow  0.808737   True   ...   0.026498
1984              Old_English_sheepdog  0.427481   True   ...   0.134269
1985                         Airedale  0.602957   True   ...   0.086276
1986                          padlock  0.731564  False   ...   0.036469
1987                       Eskimo_dog  0.482498   True   ...   0.134589
1988               miniature_pinscher  0.817066   True   ...   0.034195
1989                       dishwasher  0.700466  False   ...   0.039012
1990                       Eskimo_dog  0.616457   True   ...   0.001670
1991                 Italian_greyhound  0.734684   True   ...   0.039725
1992                       Rottweiler  0.681495   True   ...   0.024525
1993                          malamute  0.442612   True   ...   0.177822


     p3_dog  favorites  retweets  user_followers  dog_stage  rating_numerator  \
0      True        474       224         6992426       None               7.0
1      True        310        85         6992426       None              12.0
2     False        458       248         6992426       None              10.0
3     False        803       604         6992426       None               6.0
4      True        320        93         6992426       None              11.0
5      True        195        65         6992426       None               7.0
6      True        264       100         6992426       None              11.0
7     False        547       235         6992426       None               7.5
8     False        540       286         6992426       None               9.0
9      True        544       227         6992426       None              11.0
10     True       6956      2210         6992426       None              12.0
11     True        377       198         6992426       None               8.0
12     True       1745      1162         6992426       None              12.0
13     True        487       253         6992426       None              10.0
14     True        841       559         6992426       None               8.0
15     True        209        69         6992426       None               9.0
16     True        889       547         6992426       None               7.0
17    False        341       130         6992426       None               9.0
18    False       5961      2386         6993394       None               1.0
19    False        502       332         6992426       None               7.0
20     True        383       156         6992426       None              11.0
21    False        456       228         6992426       None               9.0
22    False        240        59         6992426       None               8.0
23    False        423       253         6992426       None               9.0
24     True        718       341         6992426       None               8.0
25     True        338        98         6992426       None              12.0
26     True       1084       520         6992426       None              10.0
27     True        199        58         6992426       None               8.0
```

28	True	294	65	6992426	None	12.0
29	False	375	168	6992426	None	10.0
...	...	...	...	...	...	...
1964	True	2681	807	6992420	pupper	12.0
1965	False	2046	839	6992426	pupper	10.0
1966	True	2593	759	6992421	pupper	11.0
1967	True	26058	4150	6992403	pupper	13.0
1968	True	1103	381	6992426	pupper	10.0
1969	False	11044	3583	6992419	pupper	11.0
1970	True	2391	715	6992421	pupper	10.0
1971	True	6856	1432	6992407	puppo	12.0
1972	True	15058	2243	6992403	puppo	13.0
1973	True	7939	3790	6992416	puppo	11.0
1974	True	46856	18650	6992403	puppo	13.0
1975	True	142980	48957	6992407	puppo	13.0
1976	True	5468	1438	6992416	puppo	10.0
1977	True	16672	6326	6992411	puppo	12.0
1978	True	34680	10574	6992403	puppo	13.0
1979	True	48011	10108	6992403	puppo	13.0
1980	True	19036	1640	6992403	puppo	13.0
1981	True	11130	3125	6992410	puppo	13.0
1982	True	5761	1322	6992407	puppo	13.0
1983	True	5434	1708	6992418	puppo	11.0
1984	True	3194	692	6992417	puppo	9.0
1985	True	6941	1697	6992416	puppo	10.0
1986	False	23535	8237	6993378	puppo	13.0
1987	True	9877	2927	6992412	puppo	12.0
1988	True	10954	3108	6992411	puppo	12.0
1989	True	18122	5291	6992411	puppo	12.0
1990	True	16341	2583	6992403	puppo	12.0
1991	True	19371	4155	6992403	puppo	13.0
1992	True	56057	19124	6992407	puppo	14.0
1993	True	4516	1326	6992417	puppo	10.0

	dogs_count	dog_name	dog_gender
0	1	Biden	NaN
1	1	NaN	NaN
2	1	NaN	NaN
3	1	NaN	NaN
4	1	Cupcake	NaN
5	1	Reese	NaN
6	1	Kohl	male
7	2	NaN	NaN
8	1	Philippe	NaN
9	1	Hall	NaN
10	1	Tickles	NaN
11	1	Dook	male
12	1	NaN	NaN

13	1	Naphaniel	male
14	1	Frank	NaN
15	1	Klint	NaN
16	2	Kial	NaN
17	1	NaN	male
18	1	NaN	NaN
19	1	NaN	NaN
20	1	Olive	male
21	1	Calvin	male
22	1	Jessiga	female
23	1	NaN	NaN
24	1	NaN	NaN
25	1	NaN	NaN
26	1	Filup	male
27	1	Fwed	male
28	1	NaN	NaN
29	1	NaN	NaN
...	...	...	...
1964	1	NaN	male
1965	1	Zuzu	male
1966	1	Gizmo	NaN
1967	1	Jed	male
1968	1	NaN	NaN
1969	1	Sansa	female
1970	1	Lucy	NaN
1971	1	Pablo	NaN
1972	1	Stuart	male
1973	1	NaN	NaN
1974	1	NaN	female
1975	1	NaN	NaN
1976	1	Cooper	NaN
1977	1	Lily	female
1978	1	Sebastian	male
1979	1	NaN	NaN
1980	1	NaN	female
1981	1	Tonks	female
1982	1	Oliver	male
1983	1	Bayley	female
1984	1	Abby	female
1985	1	NaN	NaN
1986	1	NaN	NaN
1987	1	Loki	NaN
1988	1	Diogi	male
1989	1	Betty	NaN
1990	1	Shikha	female
1991	1	Snoopy	NaN
1992	1	NaN	female
1993	1	Kilo	male

```
        [1994 rows x 24 columns]
```

3.1 Visualization

```
In [4]: df_master = pd.read_csv('twitter_archive_master.csv')

In [6]: df_master.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1994 entries, 0 to 1993
Data columns (total 24 columns):
tweet_id            1994 non-null int64
timestamp           1994 non-null object
source              1994 non-null object
text                1994 non-null object
expanded_urls       1994 non-null object
jpg_url             1994 non-null object
img_num             1994 non-null int64
p1                  1994 non-null object
p1_conf             1994 non-null float64
p1_dog              1994 non-null bool
p2                  1994 non-null object
p2_conf             1994 non-null float64
p2_dog              1994 non-null bool
p3                  1994 non-null object
p3_conf             1994 non-null float64
p3_dog              1994 non-null bool
favorites           1994 non-null int64
retweets            1994 non-null int64
user_followers      1994 non-null int64
dog_stage           1994 non-null object
rating_numerator    1994 non-null float64
dogs_count          1994 non-null int64
dog_name            1369 non-null object
dog_gender           862 non-null object
dtypes: bool(3), float64(4), int64(6), object(11)
memory usage: 333.1+ KB


In [10]: df_master['tweet_id'] = df_master['tweet_id'].astype(object)
         df_master['dog_stage'] = df_master['dog_stage'].astype('category')
         df_master['dog_gender'] = df_master['dog_gender'].astype('category')

         df_master.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1994 entries, 0 to 1993
Data columns (total 24 columns):
```

```
tweet_id              1994 non-null object
timestamp             1994 non-null object
source                1994 non-null object
text                  1994 non-null object
expanded_urls         1994 non-null object
jpg_url               1994 non-null object
img_num               1994 non-null int64
p1                    1994 non-null object
p1_conf               1994 non-null float64
p1_dog                1994 non-null bool
p2                    1994 non-null object
p2_conf               1994 non-null float64
p2_dog                1994 non-null bool
p3                    1994 non-null object
p3_conf               1994 non-null float64
p3_dog                1994 non-null bool
favorites             1994 non-null int64
retweets              1994 non-null int64
user_followers        1994 non-null int64
dog_stage             1994 non-null category
rating_numerator      1994 non-null float64
dogs_count            1994 non-null int64
dog_name              1369 non-null object
dog_gender             862 non-null category
dtypes: bool(3), category(2), float64(4), int64(5), object(10)
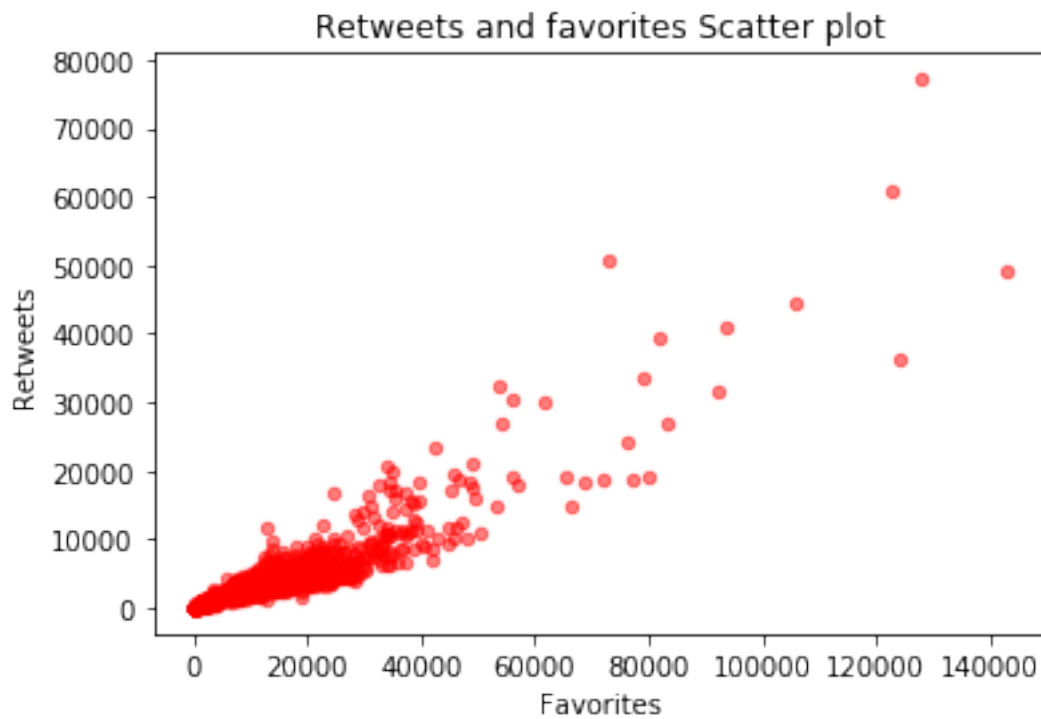memory usage: 306.1+ KB
```

```
In [11]: df_master.plot(kind = 'scatter', x = 'favorites', y = 'retweets', alpha = 0.5, color =
         plt.xlabel('Favorites')
         plt.ylabel('Retweets')
         plt.title('Retweets and favorites Scatter plot')
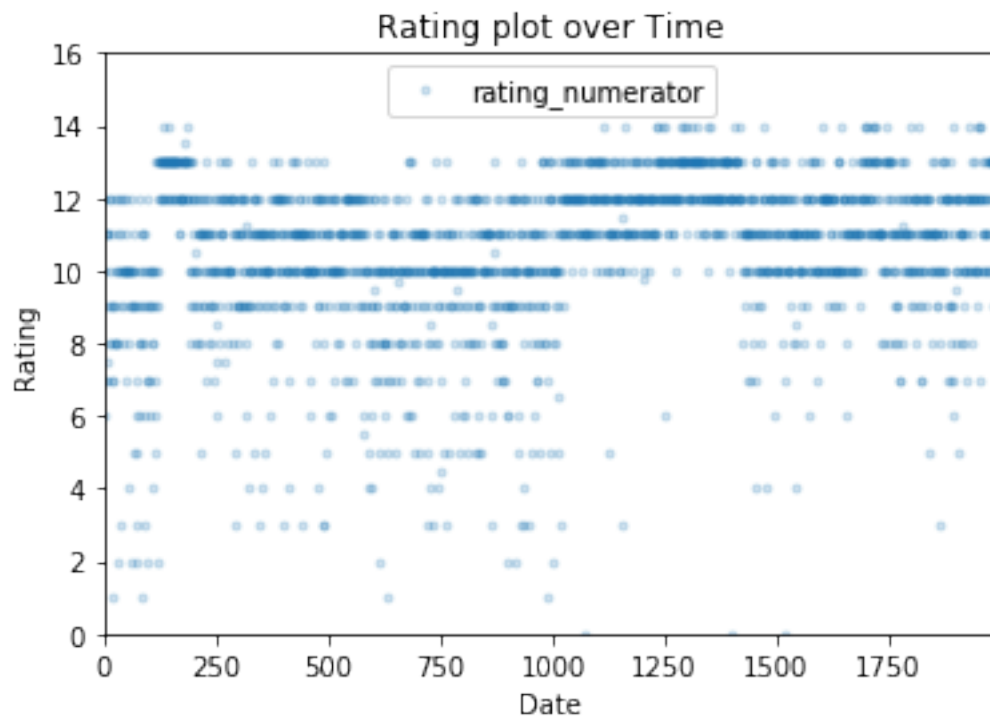
Out[11]: Text(0.5,1,'Retweets and favorites Scatter plot')
```

Retweets and favorites Scatter plot

```
In [14]: df_master.plot(y ='rating_numerator', ylim=[0,16], style = '.', alpha = .2)
         plt.title('Rating plot over Time')
         plt.xlabel('Date')
         plt.ylabel('Rating')

Out[14]: Text(0,0.5,'Rating')
```

Rating plot over Time


```
In [15]: df_master[df_master['dog_gender'].notnull()]['dog_gender'].value_counts().plot(kind = '
         plt.title('Dog Gender Partitions')
```

```
Out[15]: Text(0.5,1,'Dog Gender Partitions')
```

Dog Gender Partitions